REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER CNR 53	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER

4. TITLE *(and Subtitle)* Officer Selection Study	5. TYPE OF REPORT & PERIOD COVERED
	6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR*(s)* Peter Stoloff	8. CONTRACT OR GRANT NUMBER*(s)* N00014-76-C-0001

9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Naval Analyses 2000 No. Beauregard Street Alexandria, Virginia 22311	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS Deputy Chief of Staff (RD&S) Headquarters, Marine Corps Washington, D.C. 20380	12. REPORT DATE January 1983
	13. NUMBER OF PAGES 129

14. MONITORING AGENCY NAME & ADDRESS*(If different from Controlling Office)*	15. SECURITY CLASS. *(of this report)* Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Achievement tests, AGCT (Army General Classification Test), Aptitude tests, ASVAB (Armed Services Vocational Aptitude Battery), Intelligence tests, Marine Corps personnel, Officer personnel, Performance (human), Performance tests, Personnel selection, Predictions, Regression analysis, SAT (Scholastic Aptitude Test), Scoring, Statistical data

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

The Marine Corps uses several paper and pencil tests of mental ability as primary selection screens for officers. This study validated two of these tests, the Scholastic Aptitude Test (SAT) and the Electrical (EL) Composite of the Armed Services Vocational Aptitude Battery (ASVAB), against Officer Candidate School (OCS) and The Basic School (TBS) performance measures. The results showed both tests to be valid predictors of performance in TBS. Scores on the operational selection tests were equated with each other. The current SAT and ASVAB minimum qualifying scores were found to be about the same.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73
S/N 0102-LF-014-6601

100

OFFICER SELECTION STUDY

Peter Stoloff

CENTER FOR NAVAL ANALYSES

From: Commandant of the Marine Corps
To: Distribution List

Subj: CNA 53, "Officer Selection Study," promulgation of

Encl: (1) CNA Report 53, January 1983

1. The enclosure is the final report of the Officer Selection Study.

2. The study had two objectives:

    a. Determine the validity of the aptitude tests currently used as officer accession screens (ASVAB-EL, SAT, ACT).

    b. Determine if an improved officer selection screen can be developed from the ASVAB.

3. The study objectives were met and the enclosure is approved for distribution. The information and results therein may be used to assist further analysis as necessary.

Eugene B. RUSSELL
BRIGADIER GENERAL, U. S. MARINE CORPS
DEPUTY CHIEF OF STAFF FOR RD&S

Subj: Center for Naval Analyses Report 53

DISTRIBUTION LIST

A1	Assistant Secretary of the Navy (M&RA)
A1	Deputy Assistant Secretary of the Navy (Manpower)
A2A	Comptroller of the Navy
A2A	Office of Program Appraisal
A2A	Chief of Naval Research
A6	Deputy Chief of Staff (Manpower) HqMC
A6	Deputy Chief of Staff (Training) HqMC
	Director, Personnel Procurement Division (HqMC)
	Director, Manpower Plans & Policy Division (HqMC)
	Director, Personnel Management Division (HqMC)
A6	Deputy Chief of Staff (RD&S) HqMC
A6	Fiscal Director, HqMC
A6	Director, History and Museums, HqMC
B1B	Under Secretary of Defense for Research and Engineering
B1B	Assistant Secretary of Defense (MRA&L)
B1B	Office of Ass't Sec'y of Defense (MRA&L)(MP&FM)(AP)
B1B	Director, Program Analysis and Evaluation, OSD
B2A	Defense Technical Information Center (12 copies)
B3	National Defense University
B3	Armed Forces Staff College
FF18	Naval Tactical Support Activity
FF38	Naval Academy (Nimitz Library)
FF44	Naval War College
FF48	Naval Human Resources Management Center
FJ18	Naval Military Personnel Command
FJ76	Naval Recruiting Command
FJ89	Naval Manpower Material Analysis Center, Atlantic
FJ89	Naval Manpower Material Analysis Center, Pacific
FKA6A16	Naval Personnel R&D Center (POM Program Director, Code 12)
	Naval Personnel R&D Center (Technical Library)
FT1	Chief of Naval Education and Training
FT73	Naval Postgraduate School
FT87	Human Resources Management School
V8	Marine Corps Recruit Depot, Parris Island
V8	Marine Corps Recruit Depot, San Diego
V12	Marine Corps Development and Education Command
OpNav:	Op-09BH (Naval History)
	Op-91 (Program Resource Appraisal Division)
	Op-01 (DCNO, Manpower, Personnel & Training)
	Op-11 (Total Force Planning Division)
	Op-13 (Military Personnel and Policy Division)
	Op-15 (Human Resources Management Division)

DISTRIBUTION LIST (continued)

Other
Department of the Army (Attn: Adj Gen'l) (6 copies)
Department of the Army Library
Department of the Army Headquarters (Code DAPE-MPE-CS)
Army Research Institute (Director, Manpower & Personnel Laboratory)
Army Research Institute (Chief, Personnel Utilization Technical Area)
Army Research Institute (Technical Library)
Department of the Air Force (SAMI)
Hq Air Force Manpower & Pesonnel Center (Code MPC YPT)
Air Force Human Resources Laboratory (AFHRL/MOA)
Air Force Human Resources Laboratory (AFHRL/Technical Library)
Hq, Military Enlistment Processing Command (Code MEPCT-P)
Hq, U.S. Coast Guard (Code G-P-1/2/TP42)
Institute for Defense Analyses
Human Resources Research Organization
The Rand Corporation

OFFICER SELECTION STUDY

Peter Stoloff

ABSTRACT

The Marine Corps uses several paper
and pencil tests of mental ability as
primary selection screens for officers.
This study validated two of these tests,
the Scholastic Aptitude Test (SAT) and
the Electrical (EL) Composite of the
Armed Services Vocational Aptitude
Battery (ASVAB), against Officer Candi-
date School (OCS) and The Basic School
(TBS) performance measures. The results
showed both tests to be valid predictors
of performance in TBS. Scores on the
operational selection tests were equated
with each other. The current SAT and
ASVAB minimum qualifying scores were
found to be about the same.

EXECUTIVE SUMMARY

BACKGROUND

One method the U.S. Marine Corps uses to select officers is based on scores from paper and pencil tests of mental ability. At present, the Marine Corps uses the Scholastic Aptitude Test (SAT) or the Electrical (EL) composite of the Armed Services Vocational Aptitude Battery (ASVAB) as its primary selection screen. Current selection policy is to accept candidates with an SAT score of 1,000 or an ASVAB-EL score of 120.

This study provides a basis for evaluating existing paper and pencil tests used to select officers. In this case, the SAT and EL. It focuses on the validation of the currently used tests and the construction of a new ASVAB composite developed specifically for selecting officers. Because the Marine Corps uses alternative selection tests, the equivalency of test scores across tests becomes an important issue for setting minimum qualifying, or cut scores.

The Marine Corps administers the Army General Classification Test (AGCT) during training as the official test of record. The test has been used since 1947. It provides historical continuity for comparing the scores of Marine Corps officers. For this reason, it is important to equate scores on the various selection tests with AGCT.

METHODOLOGY

Test validation involved correlating the test scores with a limited measure of officer performance. The performance measure consisted of training grades in Officer Candidate School (OCS) and the Basic School (TBS). The primary function of OCS is to serve as a filter for TBS by screening out individuals deficient in nonacademic areas such as leadership potential and motivation. TBS instruction is oriented toward training officers in more demanding academic areas. For this reason it is to be expected that aptitude tests such as ASVAB will be more successful in predicting performance in TBS than in OCS.

A sample of officers was selected from the data contained on the Automated Systematic Recruiting Support System (ASRSS), a data processing tool used to track officer candidates recruited by an Officer Selection Officer (OSO). About 70 percent of all officer candidates selected over the last 2 years are represented. The ASRSS data covered the period from January 1981 to June 1982.

RESULTS

Validities for OCS Group

The OCS sample consisted of 2,156 males and females who had appropriate test and performance scores. The performance measure was the total OCS training grade. The grade is a composite of several components that measure physical fitness, motivation, and academic skills. However, only the total grade was available. The computed validity (correlation between test score and performance) coefficients, although positive, were not large enough to be of practical significance. Average test scores for those who completed and those who dropped the programs were about the same. We concluded that SAT and ASVAB-EL scores were not valid predictors of OCS attrition or total training grades. The lack of validity is attributed to the largely nonacademic content of OCS training.

Validities for TBS Group

The TBS sample consisted of 1,201 male and female officers. Table I shows the validities for AGCT, EL, SAT, the Air Qualification Test (AQT), the Flight Aptitude Rating (FAR), which are used as secondary screens for pilots, and the Officer Aptitude Rating (OAR), the former primary selection test for officers. The performance measure used was the TBS academic (TBS-A) grade. Also shown in the table are the correlations of the test scores with the AGCT. The validities were adjusted statistically, or corrected for range restriction, to represent the correlations between test scores and performance in the unselected reference population.

TABLE I

TEST CORRELATION COEFFICIENTS

Test	TBS-A performance	AGCT
AGCT	.75	–
ASVAB-EL	.73	.92
SAT	.69	.84
AQT	.63	.78
FAR	.37	.36
OAR	.66	.84

The validity coefficients for the primary screens (SAT and ASVAB-EL) and AGCT are uniformly high and indicate that these tests are good predictors of TBS academic performance. The high correlations of

the selection tests with AGCT indicate that SAT and ASVAB-EL seem to be measuring the same kind of abilities as AGCT.

Reliability of ASVAB-EL

The reliability of the ASVAB-EL composite was estimated by comparing scores on two forms of ASVAB that were administered to a sample of 2,005 applicants and recruits. The overall reliability of the test was estimated to be .94. We also estimated that a person with an EL of 120 (the minimum qualifying score) would be expected to be within four points of that score when retested. This result indicates that the ASVAB-EL composite has sufficiently small measurement error to be suitable for use as an officer selection instrument.

New ASVAB Officer Selection Composite

The current ASVAB composite used to select officers (EL) was designed to measure the aptitude for the enlisted Electronics Repair Military Occupational Specialty. Recently, the Marine Corps adopted it for use as an interim selector for officers to replace the OAR, which had been used for many years and whose content was believed to be compromised. Our purpose was to determine if we could construct from the ASVAB a composite that was more highly correlated with the performance measure (i.e., predictive validity), more reliable, had a greater capacity to discriminate at the high levels of ability at which officers are selected, and that measured the kinds of abilities desired in an officer. We were able to construct an ASVAB composite that was marginally better than EL. The subtests included in that composite were Arithmetic Reasoning (AR), General Science (GS), Mathematics Knowledge (MK), and Auto and Shop Information (AS). The best content of the Officer Selection Composite (OS) was determined to be a linear combination of those four subtests:

$$OS = AR + GS + MK + AS.$$

The corrected correlations of OS with TBS-A performance and AGCT were .75 and .93, respectively. The composite is similar to ASVAB-EL, with which it has three subtests in common, but contains AS rather than the Electronics Information (EI) subtest. The main advantage of using the ASVAB-OS in place of EL is that it was designed specifically for selecting officers.

Test Equating

SAT and EL scores were equated in a subset of a nationally representative sample drawn from a 1980 youth population comprising 18- to 23-year-olds who had taken both tests. The tests were equated using the equipercentile method, i.e., by determining the scores obtained by the

same cumulative proportion in the sample. An SAT score of 1,000 was found to equate to an ASVAB composite scale score of about 122.

ASVAB and SAT scores were also equated in the ASRSS sample on the basis of equal predicted performance, as measured by the TBS academic grade. The results suggested that an ASVAB score of 124 and an SAT score of 1,000 were associated with the same TBS-A grade. The levels of performance for those above the cut scores of either test were satisfactory.

The differences in the results of equating ASVAB and SAT scores with the equipercentile and performance prediction methods are attributed to certain limitations of the data. The samples available for equating SAT and ASVAB were administered the SAT up to 4 years earlier than ASVAB, potentially leading to differential inflation of ASVAB scores. This suggests that test scores obtained from high school and college students may not be directly comparable. A proper evaluation of equivalent cut scores on ASVAB and SAT that are administered several years apart should account for potential test score growth. However, we did not have the appropriate data to isolate a growth effect. Based on currently available data we estimate that an SAT score of 1,000 should equate to ASVAB scores in the range of 120 to 124.

CONCLUSIONS

- The ASVAB-EL and SAT tests are valid predictors of academic performance at TBS.

- The ASVAB-OS composite developed in this study has high validity and reliability and could replace the ASVAB-EL as a primary screen for selecting officers.

- The currently used minimum qualification scores (1,000 on SAT and 120 on ASVAB) are about the same.

- The measurement error of the ASVAB-EL composite at the cut point (EL = 120) is sufficiently small as to allow it to be used as an officer selection instrument.

RECOMMENDATIONS

- Continue to use ASVAB and SAT to select officers.

- Obtain more precise equating of SAT and ASVAB scores that would account for differences in test scores due to administering the tests at different times.

- Validate selectors against a wider spectrum of officer performance than grades in OCS and TBS.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF TABLES (Cont'd)

CHAPTER 1

INTRODUCTION

BACKGROUND

The U.S. Marine Corps selects officers based on scores of tests of mental ability. All potential officers except for participants in the Naval Reserve Officer Training Course (NROTC) program and graduates of the Naval Academy, must obtain a minimum score on one of these paper and pencil tests.

Until recently, the Officer Aptitude Rating (OAR) was the primary selection screen. However, it was believed that poor quality control of Officer Selection Officers' (OSOs) administration of this test caused a significant decline in the test's ability to screen out poor performers [1].

Evidence for the OAR's lack of validity came from its inability to predict academic success at The Basic School (TBS). The Marine Corps has used TBS grades as a surrogate measure of job performance.

To improve the quality control of testing, the Marine Corps now uses either the Electrical (EL) composite of the Armed Services Vocational Aptitude Battery (ASVAB),* the Scholastic Aptitude Test (SAT), or the American College Test (ACT) in lieu of the OAR, as alternative accession screens.**

Current selection policy is to accept candidates with an ASVAB-EL composite score of at least 120 (115 with a waiver), and an SAT score of 1,000, or an ACT score of 45, which converts to an SAT score of 1,000. However, candidates who fail to achieve the minimum SAT/ACT score can still qualify on the ASVAB. There is no formal waiver policy with respect to SAT scores.

In addition to using the SAT and ASVAB as a primary screen for officers, the Air Qualification Test (AQT) and Flight Aptitude Record (FAR) are also used as secondary screens for aviators. These tests are usually administered as part of the initial selection process as an early screen to establish a pipeline for aviators, which includes nearly half of annual officer accessions.

---

* Throughout this report, this is denoted as ASVAB-EL.
** ASVAB has been used since June 1981, and SAT since 1972.

The Marine Corps also administers the Army General Classification Test (AGCT) to newly commissioned officers upon entering TBS. This becomes the test score of record. Since 1947, AGCT scores have been collected for Marine Corps officers. These have provided historical continuity in classification testing.

PURPOSE

The purpose of this study is to provide an empirical basis for evaluating paper and pencil tests the Marine Corps uses to select officers. The study focuses on the validation of the currently used tests, the construction of a new ASVAB composite developed specifically for predicting TBS academic performance, the relationship of each of these test scores to the AGCT, and the equivalence of the various test scores.

ORGANIZATION OF REPORT

The report is divided into seven chapters. Chapter 2 discusses the reference population and several methodological issues. Chapters 3 and 4 contain the results of the validity analysis of the various selection tests using Officer Candidate School (OCS) and TBS performance as criteria. Chapter 5 describes the development of a new Officer Selection Composite and compares it to the existing selection tests. The sixth chapter shows the results of the analysis equating the various selection test scores. The final chapter presents the conclusions and recommendations.

CHAPTER 2

METHODOLOGY

As stated, the primary purpose of our research was to determine the validation of the ASVAB-EL and SAT and a new ASVAB composite as selection tests for Marine Corps officers. The term validation is used here to mean the determination of a relationship between the scores on a particular test (SAT or AGCT) and subsequent performance in OCS and TBS. The strength of the relationship is determined by the size of the correlation coefficient between the test scores and the training grades in a reference population unaffected by selection (the closer the coefficient is to one, the more valid the relationship).

Also addressed in this report is the equivalence of the scores from different tests used for selecting officers. By equivalence we mean the determination of the score values on two or more tests that represent the same level of ability or aptitude in the population from which officers are selected. Implicit in the minimum requirement of an officer candidate to have either an ASVAB-EL score of 120 or an SAT score of 1,000 for selection is the notion that these test scores represent similar levels of expected performance. This chapter then describes the methodology we used to validate and equate those test scores.

VALIDATION

When a test is used to select personnel, the resulting test score distribution will be truncated at the selection point, or cutting score. This truncation is also called range restriction. An effect of direct selection on the test, or predictor variable, is to indirectly curtail the distribution of the performance measure as well. This arises because low performance is usually associated with low test scores. Because low test scores were eliminated, the corresponding low performance scores will also be eliminated. The net result of this selection effect is to underestimate the variance of the predictor and performance variables and, hence, the correlation coefficient. This underestimation causes problems in making inferences from test results.

Because it is not practical to select a full-range sample of officers for purposes of a validation study, it becomes necessary to estimate the degree of curtailment in the sample selected and correct the correlation coefficient to represent the validity in the entire, or reference, population. This correction procedure was accomplished by estimating the test score variance in a population from which Marine Corps officer candidates are selected.

The general approach used in validating a particular selection test consisted of these steps:

- Formulate performance criterion

- Gather test scores and performance measures from restricted officer sample

- Gather test scores from unrestricted reference population

- Correlate test scores and performance measures in officer sample using test score correlations and test variances observed in reference population as basis for correcting sample correlations for range restriction.

In the remainder of this chapter we discuss the procedures and data associated with measuring performance, sampling officers' test scores, defining the reference population and obtaining their test scores, and explaining the correlation techniques used for range correction and validation. We conclude by presenting the issues and procedures followed in equating the various test scores.

Performance Criteria

Performance measurement was restricted to school performance. Traditionally, selection standards have been validated against performance in officer training courses. Although it is more desirable to validate selection instruments against post-training performance in jobs, there are several difficulties using this approach. The major difficulties have been:

- Finding a good measure of job performance

- Defining operationally what constitutes a satisfactory level of job performance

- Identifying the many factors, other than aptitude that influence job performance.

Performance measures of two varieties—attrition and final course grade—were collected for OCS and TBS. Attrition was measured dichotomously as either completion or noncompletion of the school. However, most people complete TBS, so this is not a useful measure of TBS performance.

The course grade for OCS is a composite of 17 components: seven leadership, five academic, and five physical fitness components. Only the total score was available. TBS grade is composed of 15 academic paper and pencil achievement subtests and four leadership scores. The total grade, which includes the 19 test scores, and the academic portion are reported separately.

A final course grade below 75 percent is considered a failure in either school. However, students who are observed to perform at the failing level academically will be "rolled back" into a new class. Therefore, few academic failures actually appear in the data.

OCS and TBS are sequential in nature. All officer applicants except Naval Academy graduates must complete from 10 to 16 weeks of OCS training before receiving their commissions. Immediately upon commissioning, all officers attend The Basic School for 26 weeks.

The primary function of OCS is to serve as a filter for TBS by screening out individuals deficient in nonacademic areas such as stress-coping, leadership potential, and motivation. TBS instruction is oriented toward training officers in more demanding academic areas as well as developing leadership skills. For this reason it is to be expected that aptitude tests such as ASVAB and SAT will be more successful in predicting performance in TBS than in OCS.

Validation Sample

The composition of the sample was determined by the availability of historical data. The Automated Systematic Recruiting Support System (ASRSS) maintains test scores and OCS/TBS course grades for all officer candidates recruited through an OSO. This encompasses about 70 percent of the source population. Table 1 shows the sources of Marine Corps officers and those represented on ASRSS. The major source groups not represented on ASRSS, such as Naval Academy, NROTC, and warrant officers, are not subject to the same test selection policy with which we are concerned. Therefore, the ASRSS seemed to provide the most complete and representative set of available data.

Final course grade data were obtained manually from OCS and TBS for recent graduates for two reasons. One, supplement the ASRSS data; and two, provide an indication of the level of performance achieved by the overall officer population.

The resulting data base included test score and performance information for Marines attending TBS or OCS from January 1981 through June 1982.

The majority of officers represented on ASRSS had only OAR test scores. Of the 10,524 people on ASRSS, 6,192 had OAR, 2,488 had ASVAB 8/9/10, and 3,629 had SAT test scores at the time of sampling.* Sample sizes are reported along with the results of the analysis.

---

* Current testing policy is to convert ACT into an equipercentile equivalent SAT test score. The converted ACT is recorded as an SAT score with no further reference to ACT. Thus, it was not possible to distinguish "true" and ACT-converted SAT scores.

TABLE 1

SOURCES OF MARINE CORPS OFFICERS

Commissioning program	Percent contribution[a]	Subtotal
Navy programs		20
U.S. Naval Academy	13	
NROTC	7	
Civilian sources		60
Platoon leaders class program[b]	37	
Officer candidate program[b]	20	
Women officer candidate program[b]	3	
Active duty sources		18
Enlisted commissioning program[b]	4	
Marine enlisted commissioning educational program	2	
Warrant officer program	12	
Miscellaneous sources	2	2
Total	100	100

[a]Based on about 2,000 accessions per year.
[b]Contracted by an Officer Selection Officer.

The data base was broken down into various subsets. The two major subsets contain officers with OCS and with TBS performance data. Further subdivisions were made on the basis of available test scores.

Reference Population

The initial choice of a reference population for estimating the full-range distribution of ASVAB and SAT test scores were all officer applicants tested during a recent period. This was not feasible, however, because OSOs do not retain the test score data for those not selected.

An alternate method of estimating the required test score distributions is to use available data from a population from which the majority of Marine Corps officers are drawn. Such a group would be the 18- to 23-year-old population of American youth attending or recently graduated from college. The data from the 1980 nationwide administration of ASVAB-8A, as part of the DoD sponsored Profile of American Youth study [2], was used for this purpose.

The study used a nationally representative sample designed by the National Opinion Research Center (NORC). The technical adequacy of the sample was reviewed and approved by a panel of experts on sampling techniques [3].

The officer reference population consisted of 2.5 million 18- to 23-year-old men and women who attended at least 2 years of a 4-year college. The resulting sample contained 1,065 people with the necessary ASVAB scores. Appendix A contains descriptive statistics for the variables measured in the 1980 reference population.

Although no SAT scores were collected on the 1980 sample, the procedure described in the next section allows us to estimate the unrestricted population's SAT variance, for purposes of correcting for range, the correlations of SAT with other variables.

Range Correction Procedures

The Multivariable Correction procedure was used to adjust the sample correlations among the various test scores and performance measures to what they would be in the unselected reference population of 18- to 23-year-old males and females who attended at least 2 years of a 4-year college.

The technique was first developed by Burt [4] and has been applied in several Marine Corps personnel studies ([5] and [6]). Using matrix algebra, we are able to solve for the values of the unrestricted correlations among a set of variables if we have an estimate of the unrestricted correlations and population variances for a subset of those same variables.

In the present case, the range-restricted variables consist of ASVAB-EL, SAT, OCS, and TBS performance, measured from the sample of officers on ASRSS who were selected on ASVAB and SAT. To correct the correlations among these measures to reflect their unrestricted values in the reference population, we can use the correlations among the ten ASVAB subtests we observe in both the ASRSS sample and the 1980 college reference population. Variables such as SAT, OCS, and TBS performance can be corrected because they correlate with the ASVAB variables.

Appendix A contains the correlations among the ASVAB subtests in the 1980 college reference population used for the multivariable range correction procedure. Appendix B describes the procedure in some detail and contains both the corrected and uncorrected intercorrelations among the variables from the ASRSS sample.

TEST SCORE EQUATING

The equivalence of SAT and ASVAB scores can be defined in several ways. One definition of equality can be stated in terms of predicted

performance. Those SAT and ASVAB-EL scores associated with the same level of performance can be considered equal.

A second definition of test score equality is based on comparable test score distributions in a reference population. Those SAT and ASVAB-EL scores associated with the same cumulative proportion of their respective test score distributions are considered equal. For example, if 30 percent of a reference population score 1,000 or higher on SAT and 30 percent score 120 or higher on ASVAB-EL, we consider these scores to be equivalent. This general approach of equating through cumulative test score distributions is referred to as equipercentile equating.

A distinction is sometimes made between test score equating and calibration, or scaling. The distinction is in terms of the kinds of tests used and the representativeness of the sample with respect to the reference population. An equating is done when we have parallel forms of the same test. Parallel test forms have similar content; equal means, variances, and correlations with each other; as well as uncorrelated errors of measurement.

The terms test scaling or calibration are more appropriately applied to the variety of "equating" we wish to perform with SAT and ASVAB.

Scaling is the process of ascribing the mean and standard deviation of the distribution of a reference test to some "other" test for purposes of being able to interpret these other test scores relative to the reference test. In this way the scores on the two tests are compared relative to their respective means and standard deviations. For example, a composite ASVAB score of 120 is considered to be "the same" as or equal to an SAT score of 1,000 if both scores lie an equal number of standard deviation units from their respective means in the same population.

Calibration does not require parallel forms, and may be sample specific. The characteristics of the sample determine the extent to which the results of the calibration may be generalized. Usually, the more restricted the sample, the less we can generalize.

When calibrating the scores from two tests, several conditions should be met.

A large, full-range, sample should be used. Both high and low test scores should be represented. Furthermore, the sample test score distribution should approximate that of the reference population, which we assume to be a normal distribution.

The sample should be free of selection bias. Individuals should not be excluded from the sample because they did not qualify on one or both of the tests being equated, or on a test that is highly correlated with one of those being equated. Selection bias of this sort leads to truncation and range restriction and, possibly, large errors of measurement. However, some sort of preselection almost always occurs. Therefore, we seek to minimize the bias, or ensure that it occurs to the same degree with both tests.

In addition to careful sampling procedures, certain test administration procedures should also be observed. These include a counterbalanced administration of the two tests to participants, fairly close together in time. That is, one half of the sample would take test A in the morning of a particular day and test B in the afternoon. The other half would take the tests in reverse order. One reason for administering the tests together is to preclude the problem of test score growth with time, i.e., experience, formal training.

Unfortunately, none of these conditions were met for the available sample of Marine Corps officers. The ASRSS/TBS test score data suffer from severe sampling bias, and SAT, EL, and AGCT are not all administered to the same people, or taken at the same time. While these problems can be ameliorated through range correction for purposes of test score validation, there is no comparable methodology for removing this bias to satisfy the requirements for calibrating ASVAB and SAT. For these reasons, it was not feasible to scale these test scores with the ASRSS data. Instead, a limited sample from the 1980 Youth Population who had taken both the ASVAB and SAT were used to attempt a preliminary scaling. A detailed description of the sample, calibration methods, and the rationale for their use is given along with the results in chapter 6.

Equal Performance Prediction

Another relevant definition of test score equality lies in the relationship of the test score and level of performance. We consider a particular ASVAB-EL score to be equal to the SAT score that results in the same level of expected performance. With this technique, we regress each of the two tests to be equated on a measure of performance. This

produces a linear regression equation for predicting performance from a given test score. For example:

$$\hat{Y}_{iT_1} = b_{0T_1} + b_{1T_1}X_{iT_1} \quad ,$$

where:

$$\hat{Y}_{iT_1} = \text{predicted performance associated with score } X_i \text{ on test } T_1$$

$$\hat{Y}_{iT_1} = b_{0T_2} + b_{1T_2}X_{iT_2} \quad , \text{ for test 2.}$$

Equivalent test scores can be obtained directly, using the algebraic expression:

$$X_{iT_2} = \frac{b_{0T_1} - b_{0T_2} + b_{1T1}X_{T1}}{b_{1T_2}} \quad ,$$

or graphically, by plotting the two regression equations, as shown in figure 1.
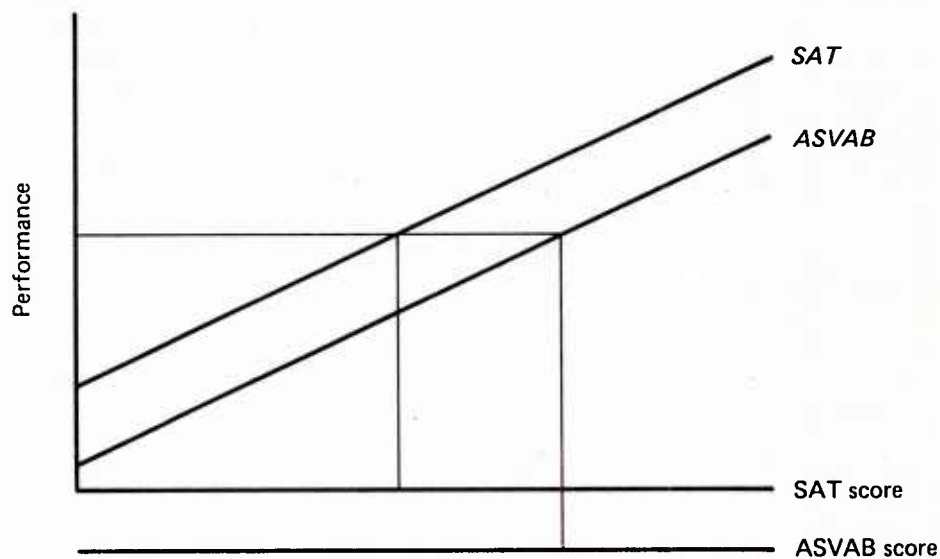
FIG. 1: EXAMPLE OF EQUIVALENT TEST SCORES

CHAPTER 3

VALIDITY OF EXISTING SELECTORS FOR OCS

In this chapter, we examine the relationship between test scores and performance in OCS. While SAT and EL are of primary interest, we also show the correlations of the AQT and FAR with the performance measure. The analysis consists of:

- Comparing the mean test scores for those who completed OCS with those who did not

- Computing the correlations of the test scores against course grade

- Correcting these correlations for range restriction.

Separate analyses were performed for each OCS training program identifiable on the ASRSS. The programs were Platoon Leader Course-Junior (PLC-Jr), PLC-combined (PLC-C), and Officer Candidate Course (OCC).* There were too few PLC-seniors with relevant test scores and performance measures to do any meaningful analysis for this group.

OCS ATTRITION

The overall rate of attrition for the 3,868 people in the sample who begin an OCS course was 24 percent. The distribution of the number of "drops" by reason is shown in table 2. The majority of drops were for medical reasons, mainly orthopedic. Only 1.7 percent of the drops, or .04 percent of those who began any class, were designated as academically unsatisfactory.

---

* The PLC programs consists of training for candidates accepted for OCS while in college. The training consists of two 6-week sessions during the summers following the junior and senior years of college and are refered to as PLC-Junior and PLC-Senior courses. The PLC-combined course is a single 10-week session following the senior year of college. The OCC program, is a single 10-week course for candidates who were selected after completing college.

TABLE 2

OCS DROPS

Category	Reason	Number of cases	Percent
Loss prior to OCS	Incompatible with Marine Corps life	115	11.0
	Financial	7	0.7
	Marriage	12	1.1
	Personal	9	0.9
	Family	11	1.0
	College dropout	3	0.3
	Other	56	5.3
Recycle	Not given	283	27.0
Medical	Dental	7	0.7
	Vision	21	2.0
	Orthopedic	175	16.7
	Blood pressure	7	0.7
	Allergy	2	0.2
	Other	152	14.5
Unsatisfactory performance in OCS	Physical fitness	12	1.1
	Leadership	160	14.3
	Academic	16	1.5
	Other	11	1.0
Total		(1,059)	(100.0)

To determine the effect of "aptitude" on attrition behavior, the mean test scores for those who dropped were compared with those who completed OCS. Because the data source does not identify the OCS program for those who didn't complete OCS, the analysis was limited to comparing test scores across programs. Also, there were too few observations to warrant separate analysis for each reason associated with the "drops."

Table 3 contains descriptive statistics for the tests of interest for those who completed and those who did not complete OCS. The results indicate that test scores are unrelated to completing OCS. Those who drop out of OCS have about the same average test scores as those who complete. The data also indicate that the mean test scores of those who complete are about the same for each program. Note also, the similarity of the performance measures across OCS classes.

TEST VALIDITIES

The multivariable range corrections were applied to the correlations and variances of the test score and performance measures of those completing OCS. The range-corrected validities are shown in table 4. The uncorrected values are contained in appendix B. The range corrections for SAT, AQT, and FAR were based on data of those who took ASVAB, in addition to SAT, AQT, or FAR, who attended a given course. Only the OCC and PLC-Jr sample sizes were large enough to warrant computing corrected validity coefficients.

The test results shown in table 4 indicate that while the correlations are positive, none is very high.

The conclusion drawn from these results is that our measure of OCS performance has modest but positive dependency on the academic abilities measured by the ASVAB-EL and SAT. As mentioned before, the OCS performance measure is probably a better indicator of leadership and physical fitness behavior, and we would not expect it to depend heavily on academic skills.

TABLE 3

TEST SCORE STATISTICS FOR THOSE COMPLETING AND THOSE NOT COMPLETING OCS COURSES

Group	Test				Course grade[a]
	ASVAB-EL	SAT	AQT	FAR	
Completed OCS					
OCC					
Mean	124	1,048	4.9	6.4	90.4
Standard deviation	7.5	156	1.2	1.5	2.9
Number of cases	262	322	249	235	716
PLC-Jr					
Mean	121	1,057	4.7	6.5	89.9
Standard deviation	6.6	144	1.1	1.5	3.6
Number of cases	184	555	542	506	1,150
PLC-C					
Mean	122	1,043	4.9	6.6	90.3
Standard deviation	6.4	137	1.2	1.7	2.8
Number of cases	39	111	135	132	290
Total completed					
Mean	123	1,052	4.8	6.4	90.0
Standard deviation	7.0	148	1.2	1.5	3.2
Number of cases	485	998	926	873	2,156
Did not complete OCS					
Mean	123	1,043	4.8	6.3	–
Standard deviation	6.8	144	1.2	1.5	–
Number of cases	338	450	401	375	–

---

[a]Grades are not assigned to those who do not complete OCS.

TABLE 4

RANGE-CORRECTED TEST VALIDITIES
AGAINST OCS GRADES[a]

Group	Test			
	ASVAB-EL	SAT	AQT	FAR
OCC	.29	.13	.32	.27
Number of cases	(262)	(105)	(102)	(92)
PLC-Jr	.21	.34	.03	.05
Number of cases	(186)	(61)	(41)	(39)

[a]Sample sizes are shown in parentheses.

CHAPTER 4

VALIDITY AND RELIABILITY OF EXISTING
SELECTORS FOR TBS

This chapter contains an investigation of the reliability of
ASVAB-EL and the validity analyses of ASVAB-EL, SAT, AGCT, as well as
AQT and FAR, against performance in TBS. Available OAR data were also
included for reference purposes. The steps in the analysis were to:

- Compute the correlation coefficients between the test
  score and performance variables

- Range correct the correlations to the reference
  population.

VALIDITY

Table 5 shows the validity coefficients after range corrections.
Range corrected and uncorrected correlation matrices among all the
variables are shown in appendix B. Descriptive statistics and bivariate
plots for selected variables are contained in appendix C.

TABLE 5

RANGE CORRECTED TEST VALIDITIES
FOR TBS PERFORMANCE

Test	Number of cases	TBS performance grade validity coefficient	
		Academic	Total
SAT	105	.72	.63
ASVAB-EL	262	.73	.56
AGCT	262	.75	.65
AQT	102	.65	.55
FAR	92	.29	.29
OAR	161	.67	.56

The resulting validity coefficients are substantial. (Even the no-longer-used OAR has a validity coefficient of .67.) As expected, the tests have higher validities against the academic grade than the total.

These validity coefficients compare favorably with those typically found between aptitude test scores and school grades. For example, Sims and Hiatt [5] reported a corrected validity coefficient of .61 between the EL composite from ASVAB, form 6/7, and training grades for Marine Corps enlisted personnel. In academic and industrial settings, correlations of .50 are considered to be high enough to justify the use of the test for selection purposes.

A useful interpretation of a validity coefficient of a test used as a selector, was developed by Brogden [7]. The interpretation is in terms of comparing the results of selection with the operational test with those of random selection (r = 0) and selection with a perfect predictor (r = 1), as extremes. The index of efficiency of the selector is the relative improvement of performance when using the selector. Operationally, the gain in expected performance will also depend on the variability of the performance measure and the selection ratio, which is the proportion of test takers we must select. With a valid test, the smaller the proportion, the higher the level of expected performance from those selected.

Expected average performance of the selected group can be expressed as

$$\tilde{Y} = \bar{Y} + \frac{h}{p} \, r \, \sigma_y \qquad (1)$$

where:

$\tilde{Y}$ = average expected performance of those selected

$\bar{Y}$ = average level of performance in the reference population

p = proportion selected (selection ratio)

h = height of normal curve corresponding to p

r = validity of selector

σ_y = standard deviation of performance measure in unselected population.

From the relationships expressed in the formula, the gain in average expected performance will be zero if the test has zero validity.

With random selection (r = 0), we would expect half of those tested to score above the mean of the reference population ($\overline{Y}$) and half below the mean. The mean expected performance level ($\overline{Y}$) of the reference population provides us with a base from which we gauge <u>relative</u> performance.

Using the results of this study, we are able to estimate all of the elements of the right side of equation (1), with the exception of Υ. We have no basis for determining the average academic performance score in the reference population of those who attended at least 2 years of a 4-year college. However, we can estimate the other elements of this equation to produce an indication of relative improvement in expected performance using ASVAB-EL and SAT as selectors.

The selection ratio for Marine Corps officers is implied by the proportion of candidates in the reference population who score at or above the minimum prerequisite score (cut score) on the selection test. The selection ratio is estimated by computing how many standard deviations the cut score is above the mean test score of the reference population. The standard score is then converted to the corresponding area under the normal curve. For ASVAB-EL, the implied selection ratio for a cut score of 120 is .34 (based on the 1980 college reference population mean of 115 and standard deviation of 11.8), while the selection ratio for those selected on SAT (cut score of 1,000) is .30 (based on the estimated college bound reference population mean of 890 and standard deviation of 206 [8]. Those numbers imply a slightly higher selection standard for SAT. The value of σ_y used in the calculations was 5.78 for ASVAB and 6.4 for SAT, which are the range corrected estimates of the standard deviation of TBS-A.

Table 6 shows the expected gain in performance over random selection, using ASVAB-EL and SAT with the validities measured in this study, and with a perfect predictor (r = 1).*

For the estimated selection ratios and amount of variability in the performance measure, average TBS-A scores could increase a maximum of about six points using ASVAB-EL and seven points using SAT. However, these tests are imperfect predictors and only a four- to five-point increase in the average performance score is realized. This average gain in predicted performance is our measure of predictive efficiency.

---

* The 4.6 point gain in TBS-A scores using ASVAB as a selector was computed from the formula, as follows:

$$4.6 = \frac{.368}{.34} (.73) (5.8) .$$

Both ASVAB-EL and SAT are seen to achieve about the same level of predictive efficiency.

TABLE 6

INCREASE IN TBS ACADEMIC GRADE SCORE
UNDER VARYING DEGREES OF TEST VALIDITY

	Random selection $(r = 0)$	Measured validities $(r = $ range corrected values$)$	Perfect prediction $(r = 1)$
SAT	0	5.3	7.4
EL	0	4.6	6.2

RELIABILITY

An important feature of a selection test is its reliability. Although there are several definitions of the term reliability, we use it to mean the consistency or "stability" of scores from the "same" test administered to the same people at different times. A perfectly reliable test would produce identical scores each time it was administered to the same person. Because there are no perfectly reliable selection tests, we ask, "How much variability is associated with a given test score?"

Test score reliability is usually reported as an average value for the test as a whole, rather than as a particular test score. However, it is important to obtain a seperate reliability estimate for scores near the cut point, or minimum qualifying score, if the average reliability estimate is not representative of these scores of interest.

The reliability of ASVAB-EL was measured empirically with the data set used to scale ASVAB form 8AX [11]. The data consisted of 2,005 pairs of EL aptitude composite scores measured with alternate forms (6 and 8AX) of the ASVAB. The sample comprised male applicants only, who were tested at Armed Forces Examining and Entrance Stations (AFEES) during 1980. This produced full range distributions of both versions of EL.

Reliability was determined for each scale score in the range of 110 to 130, the points at which selection of officers occurs. The measurement procedure was to first identify all individuals in the sample who had a particular EL score on ASVAB 8AX. We then computed the standard deviation of the corresponding ASVAB 6A EL scores, for the same people.

When computing these measures of variability, we used ASVAB 8AX scores as the reference test and ASVAB 6A as the basis for measuring the

variability. Form 6 was used because we believe examinees were more highly motivated to perform well because it was the qualifying test. This should have resulted in relatively smaller errors of measurement and, therefore, a smaller standard deviation of 6A, in comparison to 8AX scores.

The observed standard deviation of the ASVAB 6A and 8AX standard scores were 18 and 20, respectively.

To increase the number of observations the score interval for each ASVAB 8AX score was widened to $\pm$ 2 points about the score of interest. For example, the estimate of the variability in an EL score of 120 was made by computing the standard deviation of the ASVAB 6A-EL score associated with an ASVAB 8AX score of 118 through 122. Widening the score interval in this manner produced estimates similar to those obtained with an interval size of one.

Figure 2 is a graph of the ASVAB 8AX-EL standard deviations for the ASVAB 8AX-EL standard scores in the range of 110 through 130. We see that the variability is between four and five points for scores 120 and above. The data indicate that an EL score of 120 has an associated standard deviation of 4.16.
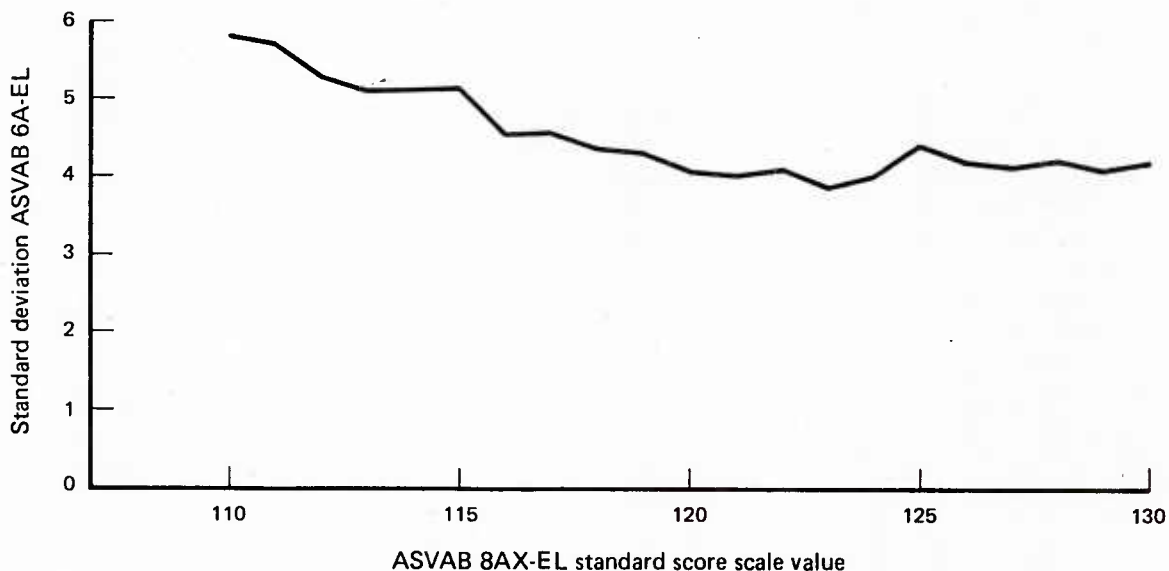
FIG. 2: VARIABILITY IN ASVAB-EL SCORES AS A FUNCTION OF SCALE VALUE

The implications of this result is that a person who scores 120 on EL on a particular administration of the test would be expected to score between 116 and 124 (120 $\pm$ 4) 68 percent of the time if he were retested. The comparable figure for a 95 percent confidence interval would be 120 $\pm$ 8. This level of precision would be equivalent to a reliability coefficient of .95.*

---

* A reliability coefficient can be computed as

$$r = 1 - \frac{se^2}{\sigma_y^2},$$

where se is the standard error of estimate, or the standard deviation of the difference between observed and expected scores, and σ_y is the standard deviation of the observed scores. When we substitute 4.16 for se and use the observed value of 17.8 for σ_y, r is equal to .95.

CHAPTER 5

ASVAB OFFICER SELECTION COMPOSITE

In this section we discuss the results of constructing several candidate ASVAB composites that could be used to select Marine Corps officers. The candidate composites are based on different philosophical approaches.

The choice of the composite for use as a selector should be governed by the following considerations:

- Validity with respect to the available short term criterion measure (TBS performance)

- Expected validity against longer-term criteria, perhaps on-the-job performance

- Capacity to measure a wide enough range of abilities, giving it the potential to measure things in common with criteria yet to be specified

- Reliability

- Information content—its capacity to discriminate at the upper end of the ability (score) scale where selection of officers is made

- Face validity—it must have a profile of subtests acceptable to the Marine Corps

- Ease of construction from the subtest standard score components—it should produce a unit weighted composite that correlates highly with the one using optimal regression weights

- Historical continuity, i.e., correlation with AGCT.

CONSTRUCTED COMPOSITES

One composite was composed of those ASVAB subtests with the highest multiple correlation with TBS academic performance. Using this as a selector should result in maximized performance, i.e., higher TBS academic grades. We labeled this the "Maximum Performance Composite."

A second composite was constructed based on the philosophy that Marine Corps officers should have a wide range of abilities (within the limits of what ASVAB measures). A preliminary factor analysis of the correlations among the ten ASVAB subtest scores from the 1980 college

reference population was done to define/describe the kinds of abilities measured. A composite we called the "Factor Composite" was then constructed from the four ASVAB subtests representing these major abilities. The composite was correlated with TBS performance to determine its validity.

A third composite was also constructed that correlated highest with AGCT and was therefore labeled the "AGCT-like Composite." Multiple correlation techniques were used for this purpose.

The ASVAB subtests for the next composites were chosen from the ASVAB subtests correlating highest with TBS performance and AGCT. Different weighting schemes were used in an effort to maximize reliability. We labeled these, "Comprehensive Composites."

The analysis used to evaluate each candidate composite consisted of:

- Determining the spectrum of abilities measured in the composite, as evidenced by its factor structure

- Comparing validity coefficients of the standard score form unweighted and regression weighted raw score form composites

- Measuring its validity against TBS academic performance

- Measuring its correlation with AGCT

- Estimating its reliability and its capability to discriminate at the higher levels of ability required of Marine Corps officers.

Two criteria for choosing a composite were not analytically evaluated. Determining the validity of the candidate composites against longer-term criteria was not attempted because of a lack of such a criterion measure. Face validity is a subjective issue. Practical and political considerations, rather than analysis, are used to gauge face validity.

In the analyses that follow, the term "regression weighted composite" is used to describe a linear composite of ASVAB subtest raw scores weighted by the empirical regression coefficients. We use the abbreviation, C_{SSS} to refer to composites that are the unweighted sums of subtest standard scores.

RELIABILITY

The methodology used to measure reliability at different ability levels made use of the data supplied in Bock and Mislevy [3] in their analysis of the data resulting from the 1980 administration of the ASVAB to the nationally representative sample of youth. These data included estimates of the standard errors of measurement for the ten ASVAB subtests at particular levels of ability. Because the Marine Corps currently selects officers with an ASVAB-EL of 120, we want high reliability at the +1 standard deviation level of ability.* Appendix D describes the method used to estimate these reliability coefficients and gives an example calculation.

The reliability estimates were made at the one standard deviation level, corresponding to a composite standard score of 120 for the 1980 DoD reference population. The test correlations reported in [9], which are also shown in appendix D, were used in the calculations.

COMPOSITE TO MAXIMIZE PREDICTED PERFORMANCE (PERFORMANCE COMPOSITE)

Table 7 shows the results of the stepwise regression analysis in support of building this composite. The raw score form of the subtests were used as the independent variables. Those ASVAB subtests that contributed significantly to the multiple correlation coefficient, above and beyond that contributed by other tests already selected by the procedure, are represented in the final set of tests. The four subtests selected by the procedure were AR, GS, MK, and AS.

TABLE 7

REGRESSION ANALYSIS OF TBS ACADEMIC PERFORMANCE
AGAINST ASVAB SUBTESTS[a]

ASVAB subtest	Standardized regression weight	F-statistic[b]	Single variable validity
AR	.36	26.3	.70
GS	.15	5.8	.59
MK	.22	11.2	.65
AS	.15	8.2	.51

[a]Number of cases was 262.
[b]Statistically significant at 1 percent level.

* The composite has a mean of 100 and standard deviation of 20 in the reference population.

The resulting composite is very similar to the EL. The new composite would substitute AS for EI, which themselves correlate .74. The multiple correlation (R) of the regression weighted raw score form composite with the TBS academic grade is .75. The SSS composite, comprising of the sum of subtest standard scores, which would be used operationally, correlates .74 with the same performance measure.

The correlation of this composite with AGCT (r_{AGCT}) is .93. The reliability of this composite for measuring ability at the 1σ level ($r_{\theta_{1\sigma}}$) is estimated to be .96.

COMPOSITE BASED ON FACTOR CONTENT

The factor analysis, although intended to provide information for building a factor "pure" composite, is also helpful for interpreting the composition of any other ASVAB based composite. The results of the factor analysis are shown in table 8.

TABLE 8

FACTOR COEFFICIENTS FOR ASVAB SUBTESTS:
COLLEGE REFERENCE POPULATION

| Test | Factor | | | | |
	I (Technical)	II (Verbal)	III (Math)	IV (Speed/ Accuracy)	Communality
GS	.63	.51	.28	.04	.80
AR	.41	.33	.70	.26	.83
WK	.29	.82	.17	.15	.77
PC	.15	.65	.20	.18	.59
NO	.05	.11	.18	.77	.62
CS	.02	.12	.08	.67	.53
AS	.86	.11	.11	.05	.80
MK	.34	.27	.70	.27	.77
MC	.74	.20	.34	.04	.77
EI	.80	.27	.21	.05	.81
Cumulative percent of variance	47	60	66	70	

Four common factors emerged from the analysis. These factors account for 70 percent of the variation among the 10 subtests. The factors measure the ability areas of:

- Technical, represented by AS, EI, MC, and GS

- Verbal, represented by WK, PC, and GS

- Math, represented by AR, MK

- Speed and Accuracy, represented by NO, CS.

The communality (h^2) indicates the proportion of common variance the test shares with the others.

The factor analysis suggests that a composite constructed of those tests with the highest correlations with each factor (the factor coefficients) would consist of AS + WK + AR + NO. Where two tests had equal coefficients on a factor, the one with the highest communality estimate was chosen.

A multiple correlation coefficient was computed for the four tests comprising this composite against TBS academic performance. Table 9 shows the results of the regression analysis used to obtain the multiple correlation coefficient. The R was .74, comparable to that obtained with the composite developed to maximize this correlation. However, the SSS composite produces an R of .70. This "shrinkage" is understandable because the composite contains a test (NO) whose marginal contribution to the valid variation in the dependent variable is small, as evidenced by the F-statistic shown in table 9. The correlation of the composite with AGCT is .90. However, this correlation would drop to .82 for the SSS composite. The reliability estimate at the 1σ ability level was .91.

AGCT-LIKE COMPOSITE

A multiple regression analysis with AGCT as the dependent variable and the ten ASVAB subtests as independent variables was performed to construct this composite. As shown in table 10, AR, MC, MK, and WK are the ASVAB tests that best predict AGCT. This combination of tests accounts for 86 percent of the variation in AGCT scores. (AR alone accounts for 77 percent.) The SSS composite would correlate .72 with the TBS academic performance measure, while AGCT itself correlates .75 with that performance measure. The reliability coefficient for the AGCT-like composite was estimated to be .96.

TABLE 9

REGRESSION ANALYSIS OF TBS ACADEMIC PERFORMANCE
AGAINST VARIABLES REPRESENTING FOUR ABILITY FACTORS

ASVAB subtest	Factor	Standardized regression weight	F-statistic[a]	Single variable validity
AR	Math	.52	84.0	.70
AS	Technical	.22	20.2	.51
NO	Speed/Accuracy	.09	4.0	.34
WK	Verbal	.08	2.4	.47

[a]All variables, except WK, statistically significant at 5 percent level.

TABLE 10

REGRESSION ANALYSIS OF AGCT SCORE
AGAINST ASVAB SUBTEST SCORES

ASVAB subtest	Standardized regression weight	F-statistic[a]	Single variable validity
AR	.49	138.4	.88
MC	.21	46.6	.71
MK	.26	46.4	.81
WK	.10	13.2	.59

[a]Statistically significant at 1 percent level.

COMPREHENSIVE COMPOSITE

The ASVAB subtests chosen for this composite consisted of the best predictors of both TBS academic performance and AGCT. The standardized regression weights presented in table 11 show the relative contribution of the tests comprising this composite for predicting TBS-A and ACGT scores. Although the composite correlates reasonably well with both AGCT and TBS-A, AR and MK consistently contribute the lion's share of the weight. The remaining ASVAB subtests contribute in different ways to the prediction of the two criteria. Whereas MC and WK add significantly and positively to AGCT, these same variables add little or negatively to TBS performance prediction. The need and justification of the higher weighting for AR and MK are apparent.

TABLE 11

STANDARDIZED REGRESSION WEIGHTS
FOR COMPREHENSIVE COMPOSITE

ASVAB subtest	Dependent variable	
	TBS-A	AGCT
AR	.37	.49
GS	.15	.05
MK	.23	.26
AS	.18	.01
MC	-.06	.19
WK	.02	.09
Multiple correlation coefficient	.75	.93

Reliabilities for the comprehensive composite were computed in two ways. Integer weights equal in value to one were used for all six ASVAB standardized subtest scores. In the second calculation, integer weights of two were used for AR and MK, while weights of one were used for the remaining ASVAB subtests. The result of double weighting the math tests in the SSS version of the composite is a slight increase in the reliability estimate, from .97 to .98.

CORRELATIONS WITH OTHER TESTS

Because any new composite is likely to be used along with SAT and AQT/FAR,* and will be referenced to AGCT, it is important to determine the relationship of the composites to these tests. Table 12 shows correlations of the various composites, including EL, with these other tests. The correlations are based on regression and unit weighted composites, comparable to what would be observed for an operational selector. The Performance composite has a nearly identical profile of correlations as EL. Note the .98 correlation of EL and the Performance composite, an indication of the near interchangeability of the two.

---

* It is recognized that the Marine Corps might like to substitute an ASVAB composite for FAR. Justification for such a substitution must be based on the validity of ASVAB with respect to an aviation-relevant performance measure. Determination of this validity is beyond the scope of this analysis. Based on available data, no ASVAB composite will produce a multiple correlation with FAR above .53. This correlation in and of itself is not large enough to warrant substitution.

TABLE 12

CORRELATIONS OF COMPOSITES WITH
THE OTHER SELECTION TESTS[a]

	Candidate composites					
Test	Performance	Factor	AGCT-like	Comprehensive[b]	Comprehesive[c]	ASVAB-EL
SAT	.79	.72	.84	.83	.82	.82
EL	.98	.86	.92	.97	.97	1.00
OAR	.86	.76	.84	.89	.89	.88
AGCT	.88	.82	.93	.88	.91	.89
AGT	.75	.72	.78	.73	.70	.79
FAR	.43	.30	.36	.44	.44	.44

[a]Unit weights, rather than regression weights, applied to subtest
standard scores.
[b]AR and MK double weighted.
[c]Equal weighted.

ABILITY AREAS MEASURED BY COMPOSITES

Table 13 groups the ASVAB tests comprising each composite by factor
content. Only the Factor composite measures all the ability areas
covered by ASVAB. The other composites are characterized by pre-
dominately measuring Quantitative and Practical Science with some Verbal
abilities. The Speed and Accuracy abilities seem to have little in
common with TBS academic performance or the other ability tests given to
Marine Corps officers.

TABLE 13

COMPARISON OF COMPOSITE FACTOR COMPOSITION

	Composite				
Content area	EL	Empirical	Factor	AGCT-like	Comprehensive
Technical	EI	AS	AS	MC	GS, MC
Verbal	GS	GS	WK	WK	WK
Math	AR	AR	AR	AR	AR
	MK	MK		MK	MK
Speed/Accuracy	–	–	NO	–	–

-29-

SUMMARY

Table 14 summarizes our findings of the feasibility of constructing a new officer selection (OS) composite with ASVAB. The sizes of the reliabilities and validities shown in table 13 suggest that any of the composites could be used for selecting Marine Corps officers. These results suggest that there would be little gained by replacing ASVAB-EL by a different ASVAB composite. However, the Performance and Comprehensive composites do show marginally greater reliability and validity estimates.

TABLE 14

SUMMARY TABLE OF VARIOUS OFFICER SELECTION (OS)
COMPOSITE EFFECTIVENESS MEASURES

	Validity of TBS-A		σ Reliability (SSS)	Correlation with AGCT	
Composite	Regression weight[a]	SSS		Regression weight[a]	SSS
EL	.74	.73	.96	.92	.89
Empirical	.75	.74	.96	.93	.88
AGCT	.73	.72	.96	.93	.91
Factor	.74	.70	.91	.90	.82
Comprehensive					
Equal weighted	.75	.72	.97	.93	.88
Math double weighted	.75	.74	.98	.93	.91

[a]Correlations based on regression weights.

Despite the good showing of the Comprehensive composite in the cited comparisons, we believe the administrative difficulties involved in using a six-variable, differentially weighted composite would make it less attractive as an operational selector.

If the Marine Corps wishes to replace EL with a different ASVAB composite, the Performance composite would be the best choice. This composite has the advantage of high reliability at the ability levels where cut scores have been set, i.e., 120 on the composite standard score scale. It would also maintain historical continuity through its high correlation with AGCT. We will refer to this composite-in-reserve as the officer selection (OS) composite, i.e., OS = AR + MK + GS + AS.

Appendix F describes the construction of the OS aptitude composite score scale and tables for converting the OS sum of subtest standard scores to the 1980 and World War II aptitude composite score scales.

CHAPTER 6

CALIBRATION OF EXISTING SELECTION TESTS

Several approaches were used to equate ASVAB-EL and SAT scores. The first was test score calibration by equipercentile equating of test score distributions in a sample. The second approach was standard score matching based on population estimates of test score parameters. The last approach involved equating test scores through equal levels of predicted performance using regression analysis. Different data sources were used with each approach.

EQUAL PERCENTILE EQUATING DATA SOURCES

Two sources of data were used for calibrating SAT and ASVAB-EL scores. The first consists of 553 pairs of SAT and AFQT scores for a subset of the 1980 youth population. The second uses the SAT population mean and standard deviation paired with the ASVAB-EL mean and standard deviation from the 1980 college reference group.

The 553 SAT scores were obtained during 1980 by the Center for Vocational Research of the Ohio State University as part of an ongoing research program. Letters were sent to the high schools of nonmilitary participants of the 1980 youth survey requesting transcripts of grades for those at least 17 years old who were no longer attending school. Although not specifically requested, 553 of the 6,591 transcripts received contained SAT scores.

We do not know how well this 553 person sample represents Marine Corps officer applicants. Although this sample is self-selected on SAT, i.e., contains only those who choose to take the SAT, it is representative of a college-bound population in terms of mean and standard deviation and contains a full range of SAT and AFQT scores.

A limiting aspect of these data is the time differential between having taken SAT and ASVAB. While SAT was generally taken by the people in this sample when they were high school seniors and 17 or 18 years old, the ASVAB scores were collected as many as 4 years later. Available data indicate that test scores increase with either increasing educational level or age. The exact relationship is not yet

established.* Unfortunately, the SAT/ASVAB data set made available for this study did not allow us to correlate test scores with either age or education. This could bias the calibration of SAT and AFQT if either test score is subject to changes during the interval. If there is a growth effect associated with increasing age or educational level, the result would be to inflate the ASVAB scores, relative to the SAT. Because of this limitation, the results of our calibration cannot be used to determine if test score prerequisites should vary with the applicant's age or education.

Another limitation of the 553 pairs of ASVAB and SAT scores was that only AFQT, rather than EL, scores were available. This is not a serious limitation in a full range sample, such as this one, because AFQT percentile scores are readily converted to composite scale scores using the standard conversion table (see appendix F) that relates an AFQT percentile score to a World War II Army standard score. This allows us to interpret the converted AFQT score as an aptitude composite. We will refer to these converted AFQT scores as ASVAB composite scale scores.

Because of the aforementioned data limitations, the results of our calibration analysis can only be approximate and should be used with allowance for this uncertainty.

POPULATION ESTIMATES OF TEST SCORE EQUIVALENCE

The second source of data used in the calibration of SAT and ASVAB-EL were estimates of the parameters (means and standard deviations) of the test score distribution in the reference population from which Marine Corps officers are selected. These data are used as a priori estimates for a common SAT/ASVAB-EL scale.

The population of 1981 SAT test takers achieved a mean of 890 with a standard deviation of 206 [8]. The 1980 youth sample consisting of those who attended at least their second year of college had a mean ASVAB-EL of 115, with a standard deviation of 11.8. These data points can be used to estimate equivalent ASVAB-EL and SAT scores if we make two assumptions. The first is that the SAT and EL scores come from equivalent groups. The second is that both measures are on an interval

---

* An analysis of the 1980 youth data [2] indicates that AFQT scores increase with age. However, this analysis did not consider amount of education. Therefore, the increase in test scores could be due to age and/or education.

scale.* If we accept these assumptions, we can compare ASVAB-EL and SAT scores by converting the scores into standard deviation (score) units and then determine the EL and SAT scores equal to the same number of standard score units. For example, an SAT score of 1,000 would be equal to +.53 standard deviation units beyond the mean, i.e., (1,000-890/206 = .53). The corresponding ASVAB-EL score, which is +.53 standard deviations beyond the EL mean, is 121. Thus, our a priori estimate is that an SAT of 1,000 should be equal to an EL score of 121. Stated in terms of an ASVAB-EL of 120, the equivalent SAT should be 977. These expected values are, within measurement error, the values of the ASVAB-EL and SAT cut scores currently used for selection.

EQUIPERCENTILE EQUATING OF SAT AND AFQT IN 1980 YOUTH SUBSET

The 553 AFQT and SAT scores available for the 1980 youth sample were equated with the equipercentile method. The equating procedure has several steps. Frequency distributions of AFQT percentile and SAT scores were constructed. The individual distributions were then smoothed using a standard smoothing package employing the cubic spline method [10]. Cumulative percentile distributions were then constructed from the smoothed frequency distributions. Appendix E contains the unsmoothed AFQT percentile score frequencies and the smoothed and unsmoothed cumulative percentile distributions. The smoothed SAT and AFQT percentile scores were then equated based on equal percentiles. AFQT percentile scores were then converted into ASVAB composite scale score form using the conversion table found in appendix F. Figure 3 shows the smoothed cumulative proportion distributions. Table 15 represents equivalent values of SAT and ASVAB composite scale scores for the limited range of scores that would be used for selection purposes. The AFQT scores are shown in composite standard score scale form to place them on the same scale as EL. These data suggest that an SAT score of 1,000 is equivalent to an ASVAB composite scale score of 122, and an ASVAB composite scale score of 120 is equivalent to an SAT score of 970 in the 1980 youth subsample. This result is in close agreement with our estimate of equivalent ASVAB and SAT scores in the college reference population.

EQUATING TEST SCORES THROUGH PREDICTED PERFORMANCE

The choice of an appropriate sample for equating SAT and ASVAB-EL through equal levels of predicted TBS-A scores was not straightforward. Because officer applicants can take ASVAB and/or SAT, three samples with

---

* An interval scale has an arbitrary origin and equally spaced units.

Therefore, a linear transformation of the forms, $\frac{x-a}{b}$, where

a is the origin and b, the scaling factor, can be used to represent a standardized value of x, independent of any particular scale.

various test score/performance measure combinations could be constructed from the ASRSS data. The preferred option is to use a sample in which each person has both test scores and the performance measure.
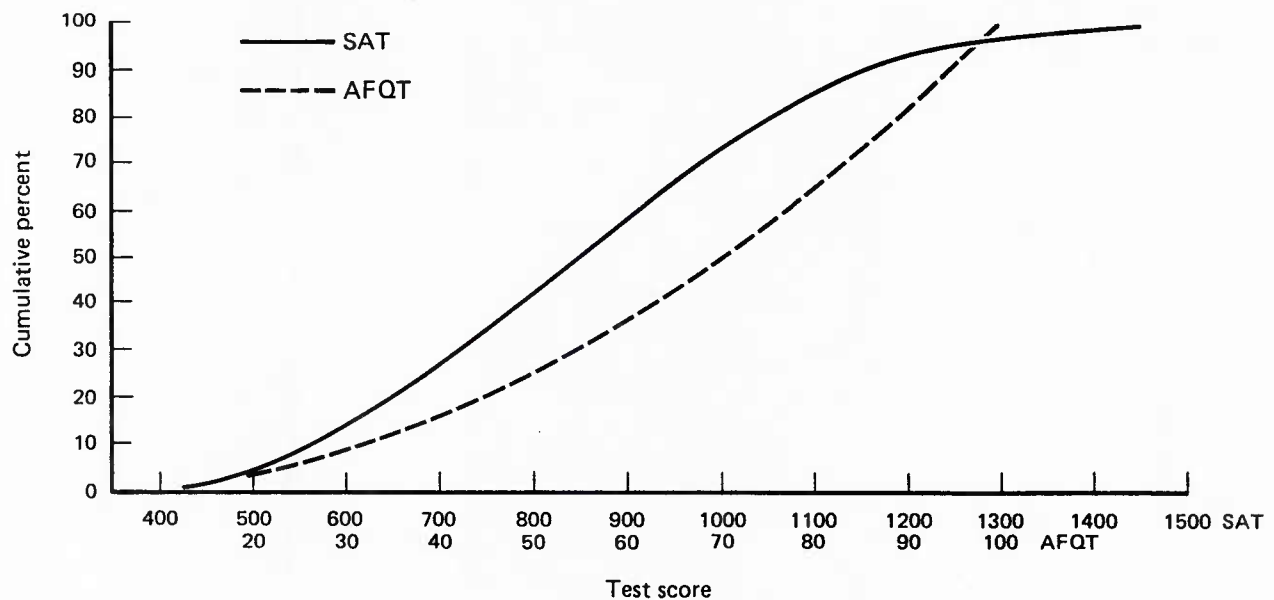
FIG. 3: CUMULATIVE PERCENTAGES OF SAT AND AFQT SCORES IN 1980 YOUTH SUBSAMPLE

A sample of 103 Marine Corps officer candidates with both ASVAB and SAT scores were used for the equating through performance. The correlation between SAT and TBS-A was .45, while the correlation between ASVAB-EL and TBS-A was .47. In raw score form, the equations for predicting TBS-A from ASVAB-EL and SAT were estimated to be

$$Y_{TBS-A} = 52 + .278 \times (ASVAB-EL) \text{ , and}$$

$$Y_{TBS-A} = 72 + .014 \times (SAT) \text{ .}$$

Both regression lines have a standard error of estimate of about four units of TBS-A. The regression lines relating SAT and ASVAB-EL scores to performance are plotted in figure 4.

TABLE 15

EQUIVALENT SAT AND ASVAB COMPOSITE SCALE SCORES
IN 1980 YOUTH SAMPLE

| Test scale		Test scale	
ASVAB	SAT	ASVAB	SAT
94	600	125	1,060
94	610	125	1,070
95	620	126	1,080
96	630	126	1,090
97	640	127	1,100
97	650	128	1,110
98	660	128	1,120
99	670	129	1,130
99	680	130	1,140
100	690	130	1,150
101	700	131	1,160
102	710	131	1,170
103	720	131	1,180
104	730	132	1,190
104	740	133	1,200
105	750	133	1,210
106	760	134	1,220
107	770	134	1,230
108	780	135	1,240
109	790	135	1,250
109	800	135	1,260
110	810	136	1,270
111	820	136	1,280
112	830	137	1,290
112	840	137	1,300
113	850	138	1,310
113	860	139	1,320
114	870	140	1,330
114	880	141	1,340
115	890	141	1,350
116	900	141	1,360
116	910	141	1,370
117	920	141	1,380
117	930	141	1,390
118	940	141	1,400
118	950	141	1,410
119	960	141	1,420
120	970	141	1,430
121	980	141	1,440
121	990	141	1,450
122	1,000	141	1,460
122	1,010	141	1,470
123	1,020	141	1,480
123	1,030	141	1,490
123	1,040	141	1,500
124	1,050		
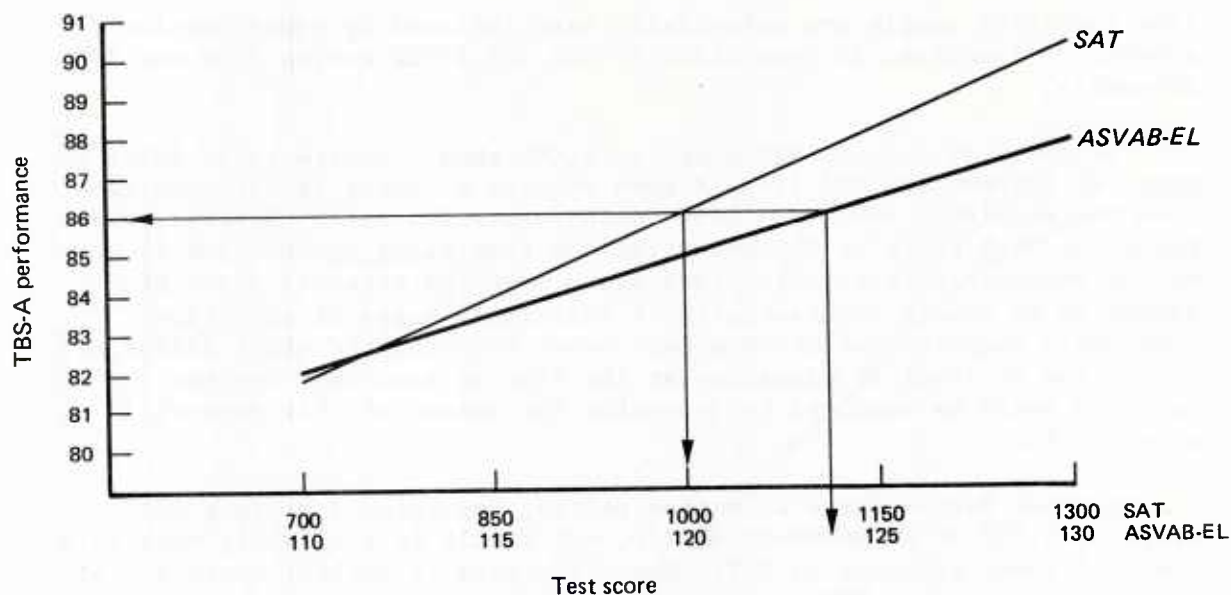
FIG. 4: TBS ACADEMIC PERFORMANCE PREDICTED FROM SAT AND ASVAB-EL SCORES (SAMPLE SIZE = 103)

These results suggest that the mean SAT score of 1,009 equates to the same level of expected performance as the mean ASVAB-EL score of 124. Conversely, an ASVAB-EL score of 120 and an SAT score of 912 also equate at the same level of expected performance.

These results appear to contradict the one obtained by the calibration methodology applied to the 553 cases from the 1980 youth population. Remember, in that case an SAT score of 1,000 equated to an ASVAB-EL of 122. The differences in the results are attributed to several factors in addition to methodology.

The composition of the two samples differed considerably. While the 1980 youth sample was an unselected (or self-selected by virtue of having taken the SAT and possessing social security numbers), full range and relatively larger sample, the ASRSS sample was highly selected and relatively small for purposes of equating. We do believe sample composition is an important factor.

Perhaps the most significant factor that would account for the different equatings, is the time differential between the administration of ASVAB and SAT. Those in the 1980 youth sample who had SAT and ASVAB scores took the ASVAB an average of 2 years after having taken the SAT. This may have resulted in some inflation of the ASVAB scores. In contrast, the 103 officers in the ASRSS sample took SAT while still in high school and ASVAB about 4 years later. Thus, the 103 ASVAB scores

REFERENCES

[1] MPI-22/1200, Decision Paper, "Officer Selection Procedures," Dec 1980

[2] Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs and Logistics). "Profile of American Youth: 1980 Nationwide Administration of the Armed Services Vocational Aptitude Battery." Mar 1982

[3] National Opinion Research Center. "The Profile of American Youth: Data Quality Analysis of the Armed Services Vocational Aptitude Battery." Chicago: NORC, Aug 1981

[4] Burt, Cyril. "Validation Tests for Personnel Selection." British Journal of Psychology 34 (1943): 1-9

[5] CNA, Study 1160, "Validation of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 6 and 7 with Applications to ASVAB 8, 9, and 10," by William H. Sims and Catherine Hiatt, Unclassified, Feb 1981

[6] CNA, Research Contribution 336, "A Method to Correct Correlation Coefficients for the Effects of Multiple Curtailment," by Thomas Mifflin and Stephen Verna, Unclassified, Aug 1977

[7] Brogden, Hubert. "On the Interpretation of the Correlation Coefficient as a Measure of Predictive Efficiency." Journal of Educational Psychology 37 (1946): 65-76

[8] College Entrance Examination Board. "National Report on College Bound Seniors, 1981." Princeton, NJ: CEEB, 1981

[9] CNA, Memorandum 82-3118, "Constructing an ASVAB Score Scale in the 1980 Reference Population," by Milton Maier and William H. Sims, Unclassified, Aug 1982

[10] Craven, Peter, and Wahba, Grace. "Smoothing Noisy Data with Spline Functions." Numerical Mathematics 31 (1979): 377-403

[11] U.S. Army Research Institute for the Behavioral and Social Science, Research Report 1301. "Scaling Armed Services Vocational Aptitude Battery (ASVAB) Form 8AX," by Milton Maier and Frances Grafton, Jan 1981

APPENDIX A

ASVAB SUBTEST MEANS, STANDARD DEVIATIONS,
AND CORRELATIONS FOR THE 1980 COLLEGE YOUTH POPULATION

APPENDIX A

ASVAB SUBTEST MEANS, STANDARD DEVIATIONS, AND CORRELATIONS FOR THE 1980 COLLEGE YOUTH POPULATION

This appendix contains tables that describe some statistics of the 1980 college youth population. Table A-1 lists the subtests that comprise the ASVAB test battery given to the 1980 youth sample and to officer candidates. Table A-2 contains the means and uncorrected standard deviations of the ten ASVAB subtests for The Basic School (TBS) and 1980 samples. The data are broken down by level of education and also show the statistics for 18- to 23-year-olds comprising the 1980 DoD reference population. Table A-3 shows the correlations among the ten ASVAB subtests for these 1,065 cases in the 1980 youth sample who attended at least 2 years of a 4-year college. They are the population correlations used to correct range restriction.

TABLE A-1

THE STRUCTURE OF ASVAB 8/9/10

Mneumonic	Subtest (content area)	Number of questions	Testing time (minutes)
GS[a]	General Science	25	11
AR[a,b]	Arithmetic Reasoning	30	36
WK[b]	Word Knowledge	35	11
PC[b]	Paragraph Comprehension	15	13
NO[b]	Numerical Operations	50	3
CS	Coding Speed	84	7
AS	Auto and Shop Information	25	11
MK[a]	Mathematics Knowledge	25	24
MC	Mechanical Comprehension	25	19
EI[a]	Electronics Information	20	9
		334	144

---

[a]These tests are used to form the Electrical (EL) composite.
[b]These tests comprise the AFQT part of the battery:

$$\text{AFQT} = \text{AR} + \text{WK} + \text{PC} + \frac{\text{NO}}{2} .$$

APPENDIX B

RESULTS OF APPLYING THE MULTIVARIABLE METHOD
TO CORRECT CORRELATION COEFFICIENTS FOR RANGE RESTRICTION

APPENDIX B

RESULTS OF APPLYING THE MULTIVARIABLE METHOD
TO CORRECT CORRELATION COEFFICIENTS FOR RANGE RESTRICTION

A major objective of this study was to estimate the correlation between selection test scores and measures of performance. But selection requirements caused sample test scores to be truncated, or restricted. Thus, the correlation matrix of all the test scores and the performance measures had to be corrected for range restriction.

The correction technique was first developed in [B-1] and has been used in numerous Marine Corps manpower and personnel studies ((B-2] and [B-3]). Our discussion of the technique parallels that given in [B-2].

The following data were needed to perform the corrections:

- Correlation matrix of all the variables (test scores and performance) in the restricted population

- Correlation matrix among the directly restricted variables in an unrestricted population

- Standard deviations of the variables contained in the aforementioned matrices.

In this study, we computed separate matrices of restricted correlations and standard deviations for the two major OCS (PLC and OCC) and TBS validation samples. However, we used a common correlation matrix among the ASVAB test score variables to represent the unrestricted correlations and standard deviations. These measures were available from the unselected 1980 youth population. We selected a sample consisting of 18- to 23-year-old males and females who attended at least 2 years of a 4-year college for this purpose. The unrestricted correlations and standard deviations for this reference group are those shown in appendix A.

Correction Equations

Suppose each member of the unrestricted population, P, is administered tests X_1, X_2, ..., X_k and a subset of the population, or restricted group, Q, is administered X_1, X_2, X_k, ..., X_{k+t}, which consists of the same k tests and t additional measures, affected by incidental selection.

We represent C_{tt} as the correlation matrix among the k + t measures taken in Q, and D_{kk} as the correlation matrix among the k measures taken in P. We would like to estimate what D_{tt}, the correlations among the full set of k + t measures in the unrestricted

population, but we are not able to measure the $k + 1, \ldots, k + t$ variables directly in P.

If we make the assumptions given next, then the correlations among the full set of $k + t$ variables in P can be estimated from C_{kt} and D_{kk}:

- The t variables can be estimated as a linear combination of the k variables (tests) subject to direct selection

- The regression lines have equal slopes in P and Q

- The errors of estimate, i.e., differences between the predicted and actual values of the t variables are the same in P and Q

- After the effects of the k explicitly selected variables are partialled out, the correlations among the t variables are the same in P and Q.

Estimation

To use the correction procedure, it is necessary to transform $\hat{D}_{kk}$ and $\hat{C}_{kt}$ into variance-covariance matrices

$$COV(X_i, X_j) = \sigma_i \sigma_j r_{ij},$$

where:

σ_i and σ_j = the standard deviations of the <u>ith</u> and <u>jth</u> variables,

r_{ij} = the correlation between i and j.

In our notation system, C will denote transformed (variance-covariance) values of $\hat{C}$, and similarly for $\hat{D}$.

We first partition C and D into submatrices that correspond to subsets of the k and t variables,

where:

C_{kk} = COV among the k tests in Q

C_{tt} = COV among the t additional measures affected by incidental selection

and similarly partition D into D_{kk} and D_{tt}.

The corrected variance-covariance submatrices are calculated as follows

$$D_{tk} = C_{tk} \, C_{kk}^{-1} \, D_{kk} \quad ,$$

and

$$D_{tt} = C_{tt} + C_{tk} \, C_{kk}^{-1}(D_{tk} - C_{tk}) \quad ,$$

where:

$$C_{tk} = \text{the transpose of } C_{kt}$$

$$D_{tk} = \text{the transpose of } D_{kt}.$$

The diagonal entries of D_{tt} will contain the range corrected variances for the t measures. The elements of D can be converted into correlations by

$$r_{ij} = D_{ij}/D_{ii} \times D_{jj} \ .$$

In the context of the current study, D_{tk} will be the corrected validity coefficients and D_{tt} the range corrected correlations among the performance measures and other variables grouped with the t variables. Because we were not able to obtain a direct estimate of the unrestricted correlations of Scholastic Aptitude Test (SAT) with the Armed Services Vocational Aptitude Battery (ASVAB) tests, we could not include SAT in the D_{kk} matrix. Instead, we incorporated SAT as one of the t variables assumed to be affected by incidental selection. This procedure will also correct for range restriction.

The range correction procedure requires that the restricted correlation matrix be based on a consistent data set. This means that to correct the correlation of SAT with TBS academic performamnce, for example, the sample correlation coefficients between the SAT score, the performance measures, as well as the ten ASVAB subtests be based on the same people. Because different people took different combinations of tests, the correction of SAT, Army General Classification Test (AGCT), ASVAB-Electrical (EL), and so on, correlations with the performance measures had to be performed independently.

By the way of illustration, the remainder of this appendix shows the results of applying the range correction procedures to ASVAB-EL and academic performance measured in the TBS sample.

Tables B-1 and B-2 show the uncorrected and corrected correlations, with standard deviations inserted in the diagonals. Note that the corrected correlations among the ASVAB tests are identical to D_{kk} used as input for the correction. For this reason we did not duplicate these values, which are already shown in appendix A.

Abbreviations Used for Variables in Data Tables

The abbreviations with an asterisk denote ASVAB raw test scores.

Mneumonic	Variable
GS*	General Science
AR*	Arithmetic Reasoning
PC*	Paragraph Comprehension
NO*	Numerical Operations
CS*	Coding Speed
AS*	Auto and Shop Information
MK*	Mathematics Knowledge
MC*	Mechanical Comprehension
EI*	Electronics Information
EL	Electrical Composite scale scores
SAT	Scholastic Aptitude Test scores
TBS-T	The Basic School total grades
TBS-A	The Basic School academic grades
AGCT	Army General Classification Test scores

TABLE B-1

UNCORRECTED CORRELATIONS FOR TBS VALIDATION SAMPLE

	EL	TBS-T	TBS-A	AGCT
GS	0.556	0.202	0.279	0.298
AR	0.611	0.314	0.378	0.643
WK	0.224	0.082	0.134	0.218
PC	0.240	0.152	0.205	0.277
NO	0.088	0.180	0.174	0.203
CS	0.025	0.137	0.124	0.175
AS	0.366	0.235	0.222	0.187
MK	0.673	0.338	0.378	0.590
MC	0.409	0.126	0.179	0.421
EI	0.644	0.149	0.199	0.301
EL	7.500	0.390	0.481	0.669
TBS-T	0.390	3.759	0.875	0.401
TBS-A	0.481	0.875	4.264	0.488
AGCT	0.669	0.401	0.488	10.217

	GS	AR	WK	PC	NO	CS	AS
GS	2.449	0.235	0.321	0.196	0.048	0.032	0.342
AR	0.235	2.878	0.180	0.167	0.218	0.113	0.099
WK	0.321	0.180	2.317	0.363	0.139	0.145	0.061
PC	0.196	0.167	0.363	1.258	0.070	0.082	-0.031
NO	0.048	0.218	0.139	0.070	5.633	0.425	-0.095
CS	0.032	0.113	0.145	0.082	0.425	13.673	-0.060
AS	0.342	0.099	0.061	-0.031	-0.095	-0.060	3.297
MK	0.257	0.463	0.182	0.257	0.293	0.177	0.135
MC	0.276	0.234	-0.029	0.101	-0.014	0.032	0.404
EI	0.296	0.213	0.091	0.081	-0.067	-0.117	0.477

	MK	MC	EI
GS	0.257	0.276	0.296
AR	0.463	0.234	0.213
WK	0.182	-0.029	0.091
PC	0.257	0.101	0.081
NO	0.293	-0.014	-0.067
CS	0.177	0.032	-0.117
AS	0.135	0.404	0.477
MK	3.103	0.272	0.182
MC	0.272	3.213	0.371
EI	0.182	0.371	2.107

RANGE CORRECTED CORRELATIONS FOR
TBS VALIDATION SAMPLE

	EL	TBS-T	TBS-A	AGCT
GS	0.858	0.489	0.594	0.680
AR	0.873	0.621	0.703	0.884
WK	0.642	0.368	0.467	0.593
PC	0.495	0.367	0.440	0.501
NO	0.310	0.339	0.336	0.352
CS	0.232	0.256	0.253	0.284
AS	0.660	0.478	0.507	0.539
MK	0.841	0.589	0.654	0.812
MC	0.745	0.450	0.532	0.706
EI	0.821	0.458	0.540	0.641
EL	11.800	0.635	0.734	0.889
TBS-T	0.635	5.037	0.917	0.655
TBS-A	0.734	0.917	6.442	0.748
AGCT	0.889	0.655	0.748	20.489

REFERENCES

[B-1] Burt, Cyril. "Validation Tests for Personnel Selection."
       British Journal of Psychology 34 (1943): 1-9

[B-2] CNA, Research Contribution 336, "A Method to Correct Correlation
       Coefficients for the Effects of Multiple Curtailment," by Thomas
       Mifflin and Stephen Verna, Unclassified, Aug 1977

[B-3] CNA, Study 1160, "Validation of the Armed Services Vocational
       Aptitude Battery (ASVAB) Forms 6 and 7 with Applications to ASVAB
       8, 9, and 10," by William H. Sims and Catherine Hiatt,
       Unclassified, Feb 1981

APPENDIX C

STATISTICAL SUMMARY OF TBS SAMPLE DATA

APPENDIX C

STATISTICAL SUMMARY OF TBS SAMPLE DATA

This appendix presents data on The Basic School in tabular and graphic forms. Tables C-1 through C-15 contain frequency distributions of the TBS variables.

Figures C-1 through C-8 are scatter plots for selected pairs of variables used in the TBS validity analysis.

The mneumonics used for variable names are the same as those used in appendices A and B.

TABLE C-1

ELECTRICAL COMPOSITE DISTRIBUTION
FOR TBS SAMPLE OF 262

Value	Frequency
98	1
106	3
109	1
110	3
113	2
115	2
116	6
117	12
118	8
119	14
120	15
121	16
122	16
123	32
124	23
125	1
126	21
127	10
128	9
129	6
130	3
131	7
132	2
133	7
134	5
135	2
137	7
141	2
145	2
147	2
150	1
153	1
155	1

Note: Sample mean is 124
       Standard deviation is 7.5.

TABLE C-2

MECHANICAL COMPREHENSION DISTRIBUTION
FOR TBS SAMPLE OF 262

Value	Frequency
9	1
10	3
11	1
13	3
14	4
15	10
16	16
17	14
18	19
19	26
20	27
21	32
22	38
23	31
24	25
25	13

Note: Sample mean is 20.2
        Standard deviation is 3.2.

TABLE C-3

GENERAL SCIENCE DISTRIBUTION
FOR TBS SAMPLE OF 262

Value	Frequency
10	1
14	3
15	3
16	2
17	3
18	9
19	22
20	26
21	27
22	45
23	48
24	46
25	28

Note: Sample mean is 21.9
      Standard deviation is 2.4.

TABLE C-4

PARAGRAPH COMPREHENSION DISTRIBUTION
FOR TBS SAMPLE OF 262

Value	Frequency
6	1
10	3
11	8
12	24
13	41
14	88
15	96

Note: Sample mean is 13.9
Standard deviation is 1.3.

TABLE C-5

WORD KNOWLEDGE DISTRIBUTION
FOR TBS SAMPLE OF 263

Value	Frequency
13	1
24	2
25	1
27	1
28	3
29	6
30	10
31	14
32	20
33	39
34	76
35	90

Note: Sample mean is 33.3
Standard deviation is 2.3.

TABLE C-6

ARITHMETIC REASONING DISTRIBUTION
FOR TBS SAMPLE OF 262

Value	Frequency
14	1
17	1
19	5
20	4
21	3
22	9
23	10
24	9
25	20
26	22
27	36
28	39
29	53
30	50

Note: Sample mean is 27.0
Standard deviation is 2.9.

TABLE C-7

NUMERICAL OPERATIONS DISTRIBUTION
FOR TBS SAMPLE OF 263

Value	Frequency
16	1
20	1
26	1
27	1
28	1
29	1
32	2
33	2
34	2
35	5
36	1
37	7
38	6
39	6
40	8
41	9
42	13
43	8
44	14
45	11
46	13
47	15
48	21
49	41
50	73

Note: Sample mean is 45.4
       Standard deviation is 5.6.

TABLE C-8

ELECTRONICS INFORMATION DISTRIBUTION
FOR TBS SAMPLE OF 263

Value	Frequency
10	1
11	6
12	10
13	16
14	21
15	46
16	49
17	49
18	32
19	23
20	10

Note: Sample mean is 16.0
Standard deviation is 2.1.

TABLE C-9

CODING SPEED DISTRIBUTION
FOR TBS SAMPLE OF 262

Value	Frequency
8	1
19	1
20	2
25	1
27	2
28	3
29	2
32	1
33	1
34	1
35	1
36	2
37	2
39	1
40	3
41	1
42	3
43	1
44	3
45	3
46	6
47	2
48	8
49	2
50	6
51	9
52	5
53	5
54	8
55	10
56	11
57	8
58	6
59	6
60	14
61	6
62	7
63	12
64	7
65	9
66	7

TABLE C-9 (Cont'd)

Value	Frequency
67	7
68	4
69	9
70	8
71	2
72	7
73	4
75	5
76	2
77	3
78	2
79	1
80	2
81	5
82	3
83	2
84	7

Note: Sample mean is 58.8
        Standard deviation is 13.7.

TABLE C-10

AUTO AND SHOP INFORMATION DISTRIBUTION
FOR TBS SAMPLE OF 262

Value	Frequency
10	2
11	1
12	4
13	2
14	8
15	10
16	8
17	19
18	18
19	25
20	27
21	34
22	23
23	37
24	31
25	13

Note: Sample mean is 20.2
Standard deviation 3.3.

TABLE C-11

MATHEMATICS KNOWLEDGE DISTRIBUTION
FOR TBS SAMPLE OF 262

Value	Frequency
6	1
11	1
12	3
13	2
14	4
15	2
16	9
17	10
18	10
19	24
20	26
21	37
22	35
23	36
24	31
25	31

Note: Sample mean is 21.1
      Standard deviation is 3.1.

TABLE C-12

ARMY GENERAL CLASSIFICATION TEST
DISTRIBUTION FOR TBS SAMPLE OF 1,200

Value	Frequency
89	1
91	1
97	1
99	6
101	1
103	7
105	9
107	13
109	16
111	31
113	24
115	30
117	48
119	75
121	79
123	98
125	100
127	86
129	129
131	81
133	53
135	73
137	56
139	50
141	37
143	20
145	18
147	11
149	26
151	9
153	2
155	7
157	1
171	1

Note: Sample mean is 128.0
       Standard deviation is 10.4.

TABLE C-13

TBS–ACADEMIC PERFORMANCE
DISTRIBUTION FOR TBS SAMPLE OF 1,201

Value	Frequency
71	1
74	1
75	6
76	12
77	8
78	24
79	31
80	52
81	62
82	81
83	72
84	77
85	111
86	109
87	107
88	98
89	73
90	78
91	64
92	54
93	38
94	25
95	16
97	1

Note: Sample mean is 86.0
Standard deviation is 4.3.

TABLE C-14

TBS-TOTAL PERFORMANCE
DISTRIBUTION FOR TBS SAMPLE OF 1,201

Value	Frequency
74	1
75	2
76	2
77	9
78	24
79	29
80	46
81	62
82	76
83	105
84	111
85	133
86	109
87	125
88	104
89	75
90	62
91	42
92	38
93	28
94	10
95	7
97	1

Note: Sample mean is 95.6
       Standard deviation is 3.8.

TABLE C-15

SCHOLASTIC APTITUDE TEST
DISTRIBUTION FOR TBS SAMPLE OF 318

Score interval	Frequency	Score interval	Frequency
600-619	1	1,060	9
620	2	1,080	14
640	0	1,100	20
660	1	1,120	12
680	0	1,140	20
700	1	1,160	9
720	2	1,180	11
740	5	1,200	10
760	0	1,220	6
780	7	1,240	10
800	5	1,260	4
820	5	1,280	1
840	7	1,300	5
860	9	1,320	4
880	13	1,340	0
900	9	1,360	3
920	6	1,380	0
940	15	1,400	1
960	14	1,420	0
980	10	1,440	1
1,000	26	1,460	2
1,020	14	1,480	0
1,040	19	1,500-1,519	1

Note: Sample mean is 1,043.
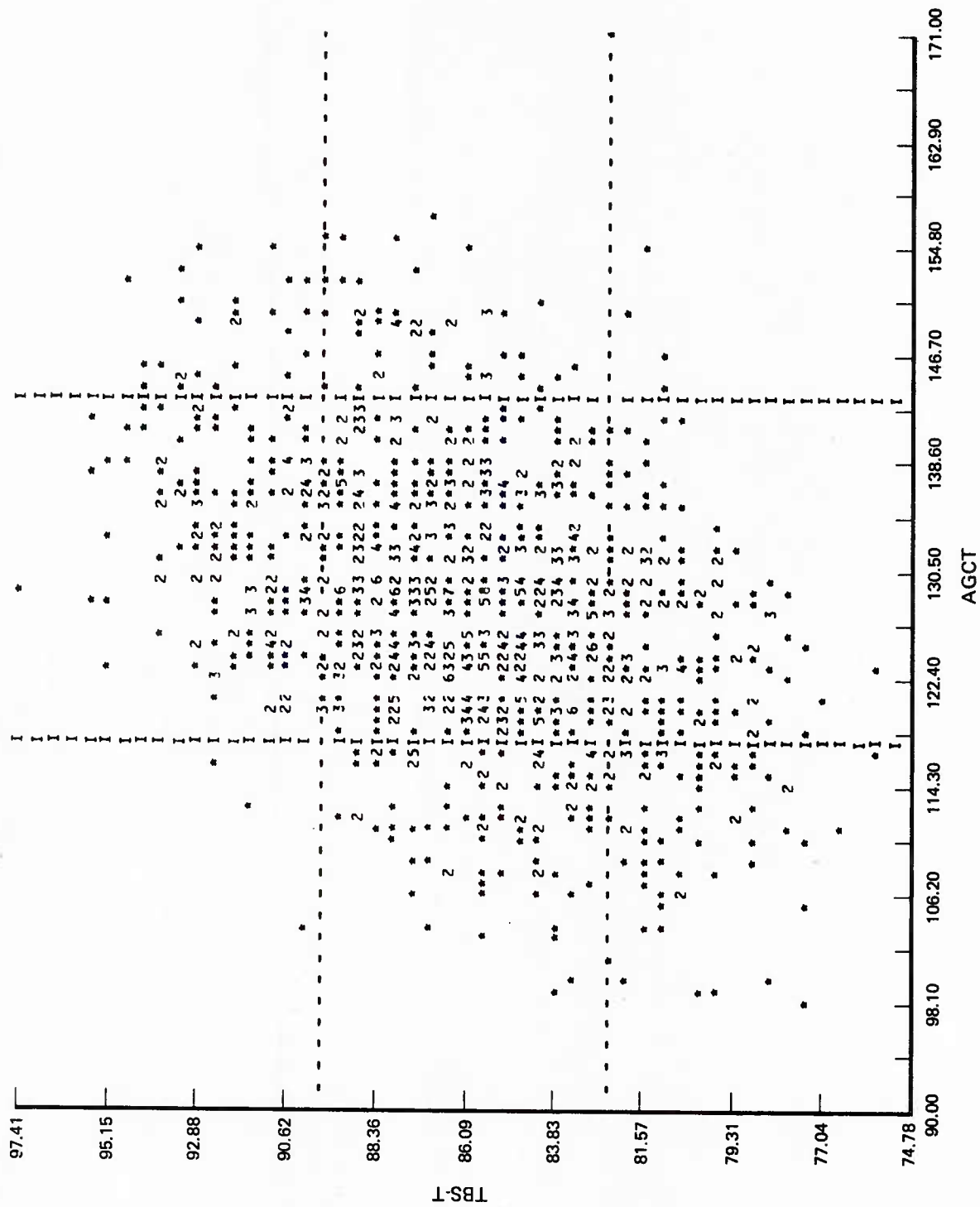   Standard deviation is 157.
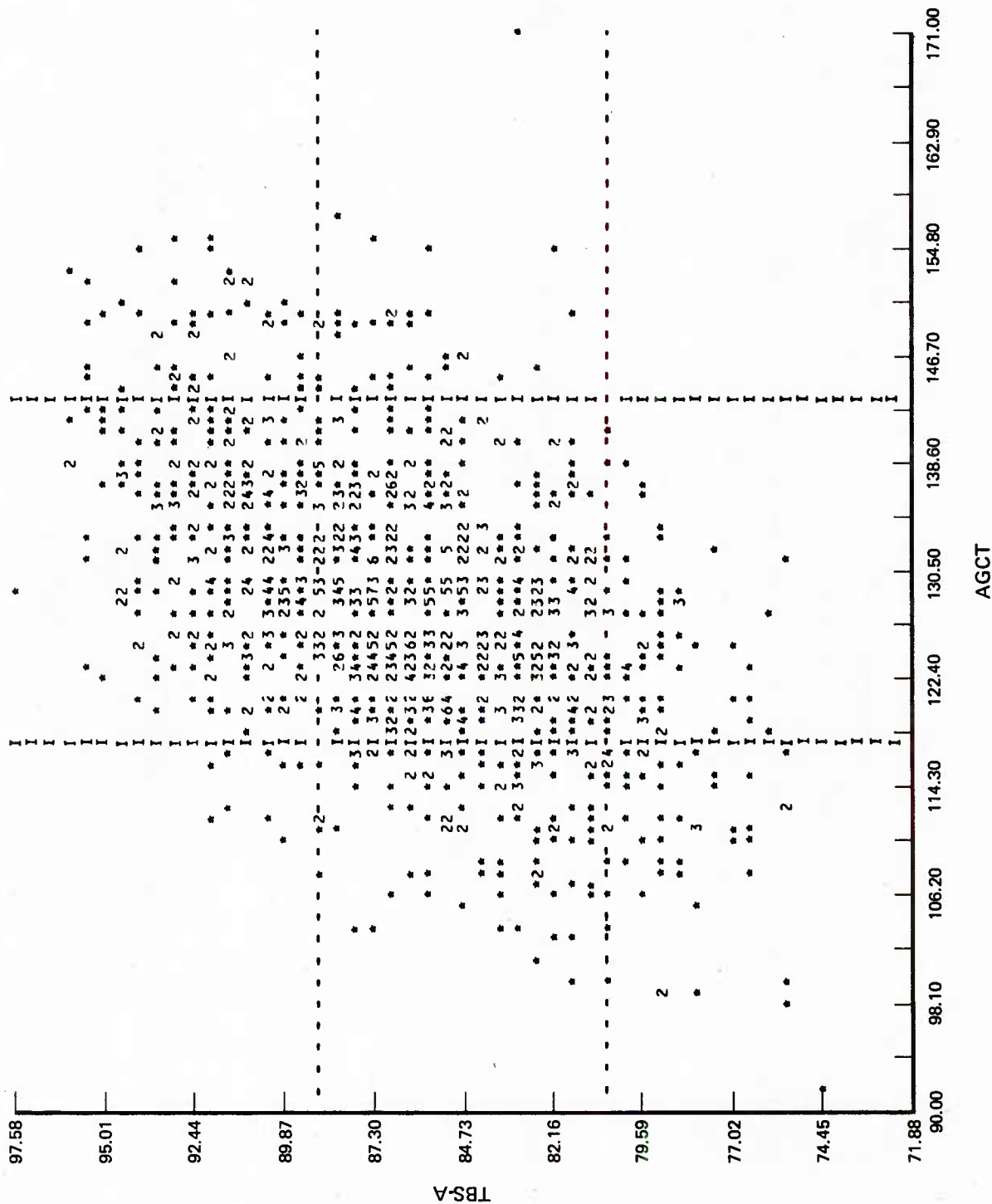
FIG. C-1: SCATTERGRAM OF TBS-T AND AGCT

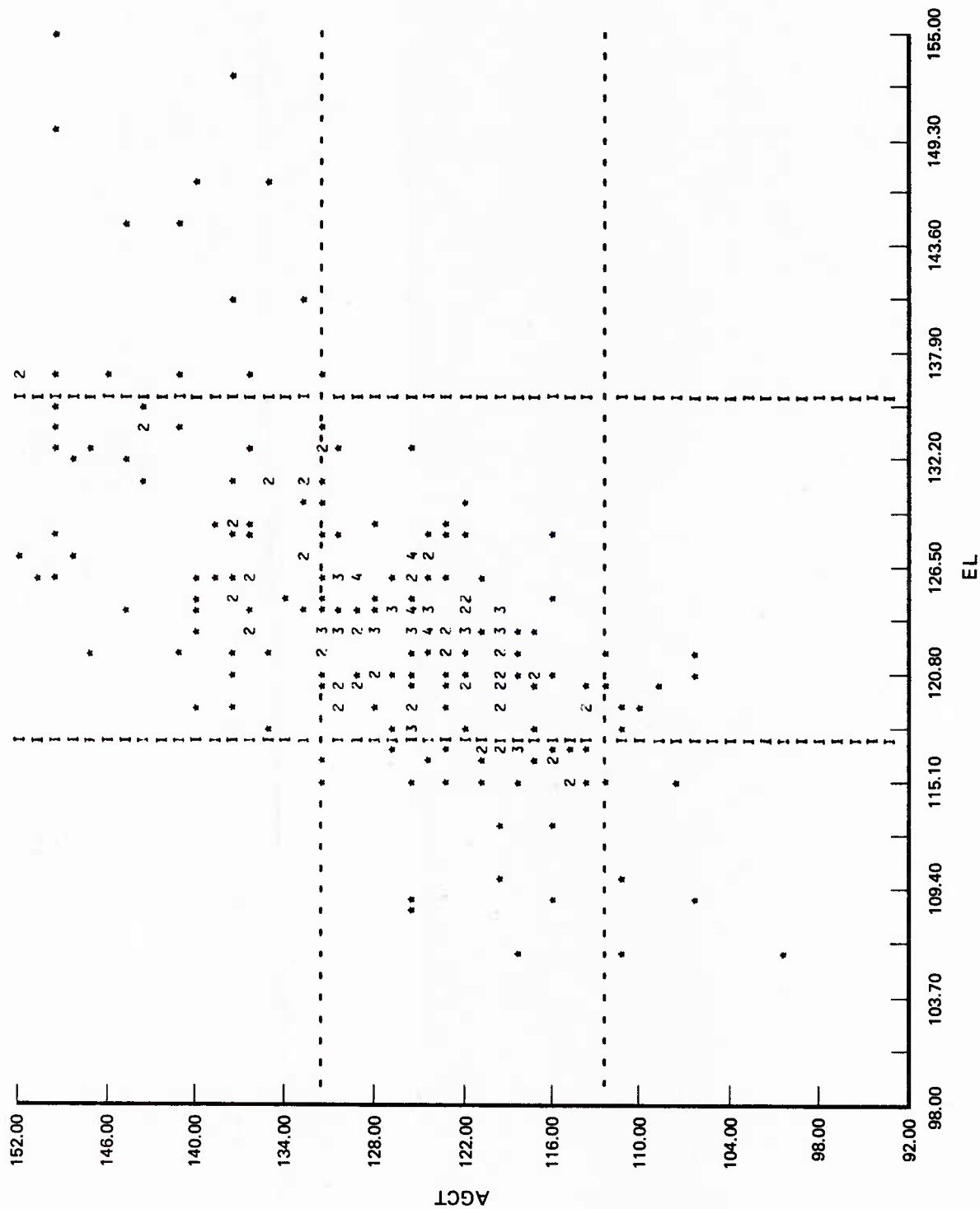FIG. C-2: SCATTERGRAM OF TBS-A AND AGCT
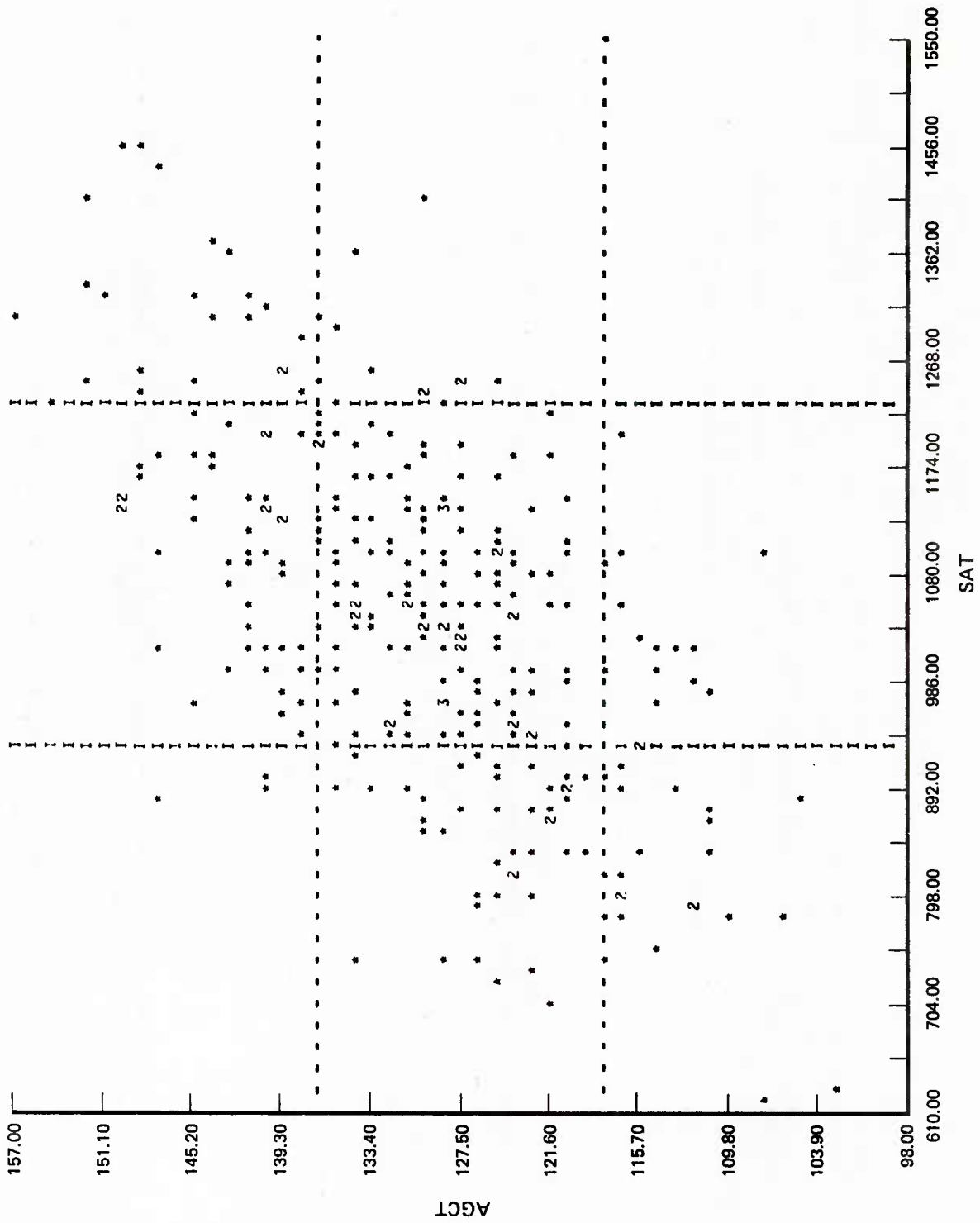
C-18

FIG. C-3: SCATTERGRAM OF AGCT AND EL

FIG. C-4: SCATTERGRAM OF AGCT AND SAT

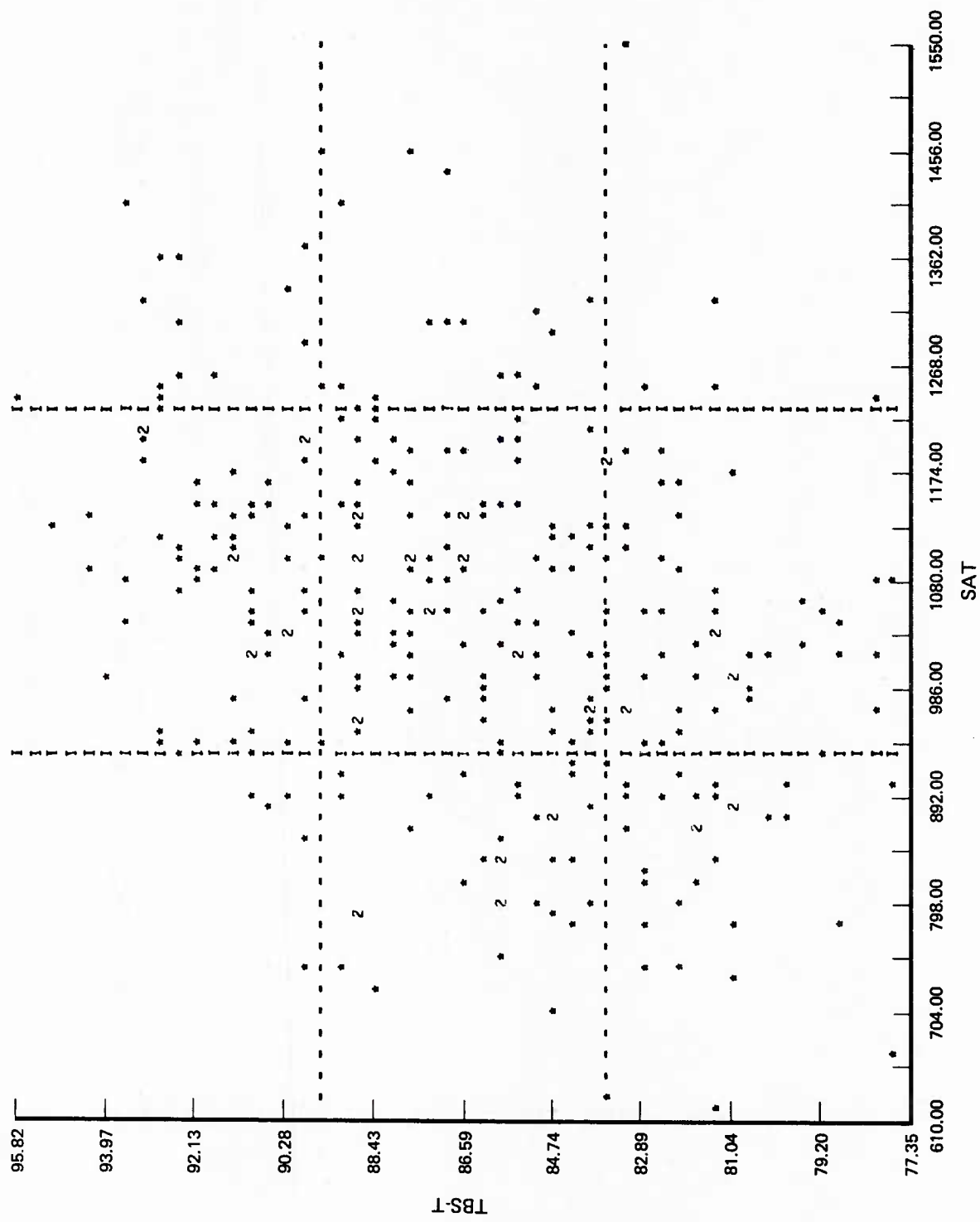FIG. C-5: SCATTERGRAM OF TBS-T AND SAT
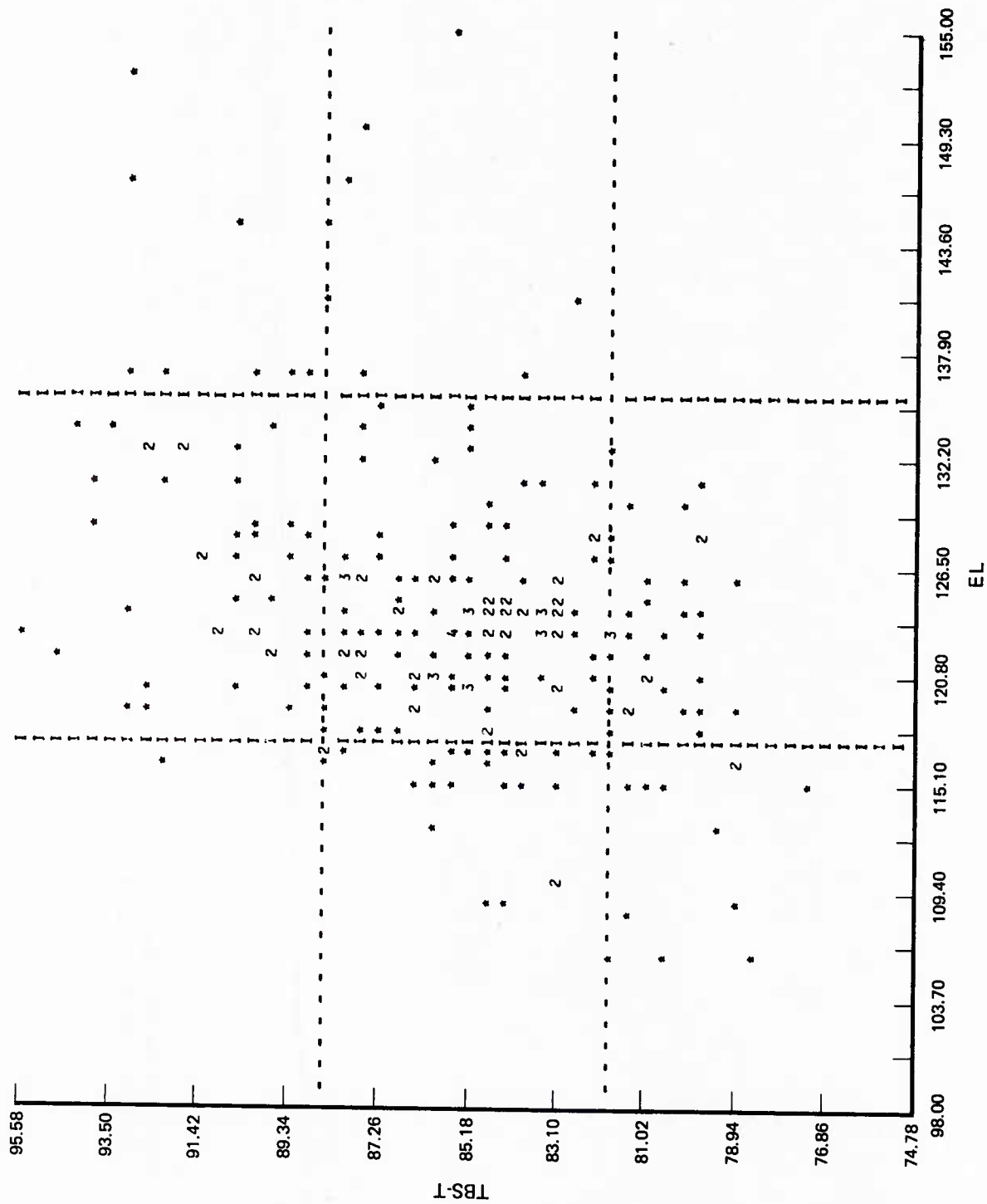
FIG. C-6: SCATTERGRAM OF TBS-A AND SAT
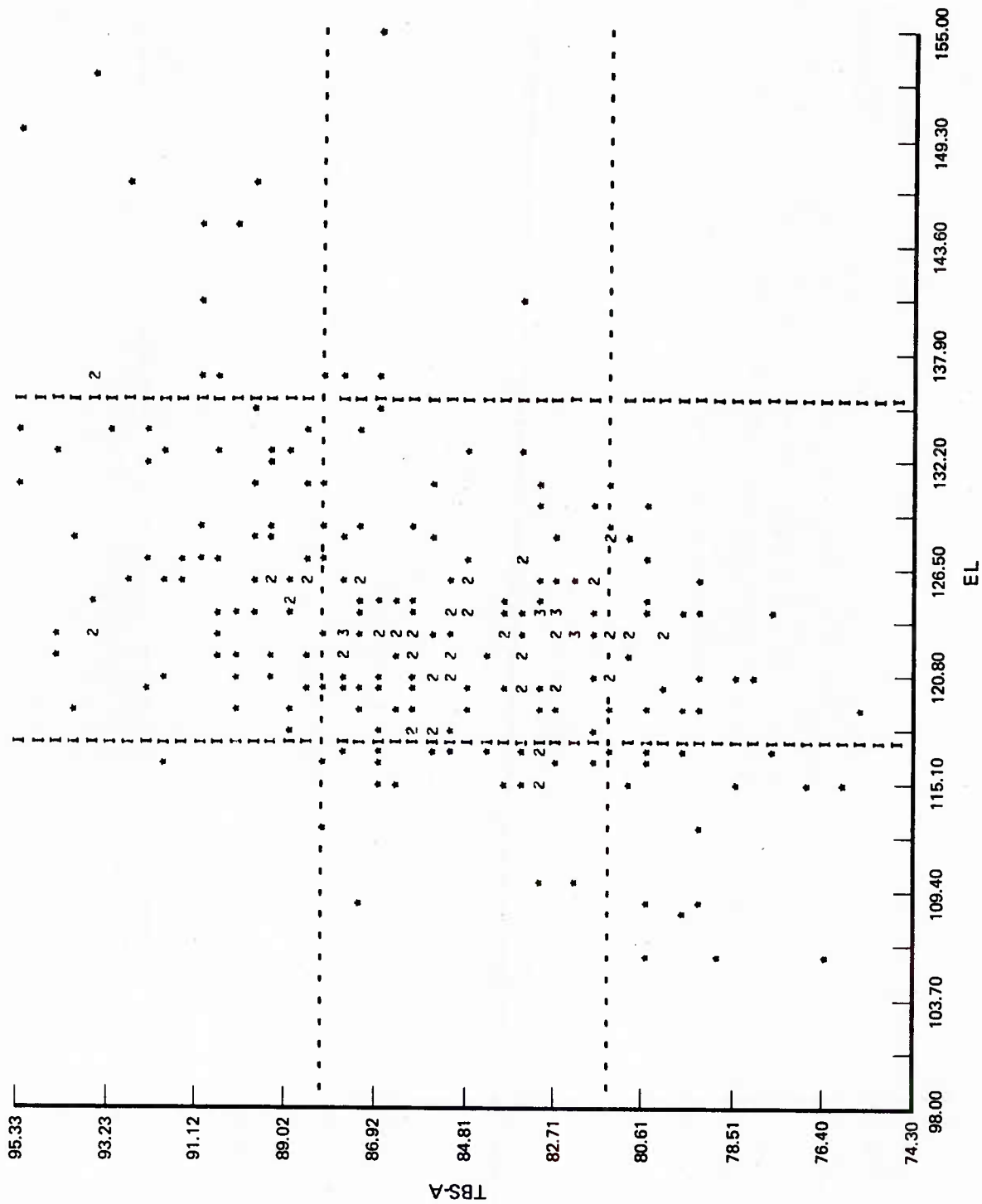
FIG. C-7: SCATTERGRAM OF TBS-T AND EL

FIG. C-8: SCATTERGRAM OF TBS-A AND EL

APPENDIX D

CALCULATION OF COMPOSITE RELIABILITY

APPENDIX D

CALCULATION OF COMPOSITE RELIABILITY

The reliabilities of the composites for candidate officer selection were estimated according to the method used to combine reliabilities of tests in a battery [D-1]. The composite reliability coefficient, R_c, is given as

$$R_c = 1 - \frac{\sum se_i^2}{\sum R}, \qquad (D-1)$$

where:

R = the full matrix of correlations among the subtests in the composite (table D-1)

se_i = the standard error of measurement of the ith subtest in standard score form.

Reference [D-2] reports the standard errors of measurement at particular levels of ability (θ) in the 1980 DoD reference population for the ten ASVAB subtests. These are reproduced in table D-2. The standard error of measurement is related to the reliability coefficient of the test as follows:

$$r_\theta = 1 - se_\theta^2 \quad .$$

The computation of R_c is shown in table D-3.

TABLE D-1

CORRELATIONS[a] OF ASVAB-8 SUBTESTS
FOR MALES AND FEMALES IN THE 1980 REFERENCE POPULATION[b]

	ASVAB subtests										Mean	Standard deviation
	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI		
GS	–	72	80	69	52	45	64	69	70	76	16.0	5.01
AR	72	–	71	67	63	51	53	83	69	66	18.0	7.37
WK	80	71	–	80	60	55	53	67	60	68	26.3	7.71
PC	69	67	80	–	60	56	42	64	52	57	11.0	3.36
NO	52	63	60	60	–	70	30	62	40	41	34.5	10.99
CS	45	51	55	56	70	–	22	52	34	34	46.3	16.25
AS	64	53	53	42	30	22	–	41	74	75	14.3	5.55
MK	69	83	67	64	62	52	41	–	60	59	13.6	6.39
MC	70	69	60	52	40	34	74	60	–	74	14.2	5.35
EI	76	66	68	57	41	34	75	59	74	–	11.6	4.24

---

[a]Decimals omitted.
[b]From CNA, Memorandum 82-3118, "Constructing an ASVAB Score Scale in the 1980 Reference Population," by Milton Maier and William H. Sims, Unclassified, Aug 1982
GS = General Science
AR = Arithmetic Reasoning
WK = Word Knowledge
PC = Paragraph Comprehension
NO = Numerical Operations
CS = Coding Speed
AS = Auto/Shop Information
MK = Mathematics Knowledge
MC = Mechanical Comprehension
EI = Electronics Information.

TABLE D-2

SUBTEST STANDARD ERRORS OF MEASUREMENT

Subtest	Items	Standard deviations from mean				
		-2	-1	0	+1	+2
General Science	25	.59	.31	.33	.39	.67
Arithmetic Reasoning	30	.83	.33	.23	.31	.59
Word Knowledge	35	.37	.20	.23	.44	.94
Paragraph Comprehension	15	.69	.33	.45	.78	1.56
Numerical Operations*	50	.39	.52	.64	.73	.77
Coding Speed*	84	.25	.37	.47	.56	.61
Auto and Shop Information	25	.96	.50	.25	.31	.72
Mathematics Knowledge	25	.89	.20	.25	.29	.61
Mechanical Comprehension	25	.82	.42	.33	.42	.65
Electronics Information	20	.94	.42	.35	.47	.75

Source: Reference D-2.
Note: (*) indicates speeded subtests. Subtest scores standardized to have zero mean and unit variance.

TABLE D-3

EXAMPLE CALCULATION OF R_c FOR AR + GS + MK + AS

	Correlations				
	AR	GS	MK	AS	se[a]
AR	1.00	.72	.83	.53	.31
GS	.72	1.00	.69	.64	.39
MK	.83	.69	1.00	.41	.29
AS	.53	.64	.41	1.00	.31

Note: $R_c = 1 - \dfrac{.428}{11.64} = .96$.

[a]Standard error of measurement at one standard deviation from the mean.

REFERENCES

[D-1] Horst, Paul. <u>Psychological Measurement and Prediction.</u> Belmont, CA: Wadsworth, 1966

[D-2] National Opinion Research Center "The Profile of American Youth: Data Quality Analysis of the Armed Services Vocational Aptitude Battery," Chicago, Aug 1981

APPENDIX E

AFQT AND SAT SCORE DISTRIBUTIONS
FROM 1980 YOUTH SUBSAMPLE

APPENDIX E

AFQT AND SAT SCORE DISTRIBUTIONS
FROM 1980 YOUTH SUBSAMPLE

Cumulative score distributions for SAT and ASVAB AFQT were gener-
ated from the 553 score pairs available from the 1980 youth sample. The
distributions are shown in tables E-1 and E-2. The AFQT frequencies are
based on the number of observations in each score interval. SAT fre-
quencies are based on score intervals of 10. The cumulative proportions
shown are the result of accumulating the observed as well as the
smoothed frequencies. Smoothing was done using the cubic spline method,
which fits a third-degree least squares equation through the observed
frequencies in nonaccumulated form.

CUMULATIVE DISTRIBUTION OF AFQT 8/9/10 SCORES
IN 1980 YOUTH SUBSAMPLE

AFQT percentile	Unsmoothed frequency	Cumulative proportion	
		Unsmoothed frequency	Smoothed frequency
1.0	1	.0018	.0000
2.0	0	.0018	.0002
3.0	0	.0018	.0006
4.0	0	.0018	.0012
5.0	0	.0018	.0021
6.0	1	.0036	.0031
7.0	0	.0036	.0043
8.0	1	.0054	.0058
9.0	3	.0108	.0074
10.0	2	.0145	.0093
11.0	1	.0163	.0113
12.0	1	.0181	.0136
13.0	1	.0199	.0161
14.0	1	.0217	.0187
15.0	4	.0289	.0216
16.0	4	.0362	.0247
17.0	2	.0398	.0280
18.0	1	.0416	.0315
19.0	1	.0434	.0352
20.0	4	.0506	.0391
21.0	4	.0579	.0433
22.0	3	.0633	.0476
23.0	1	.0651	.0521
24.0	4	.0723	.0569
25.0	6	.0832	.0618
26.0	3	.0886	.0670
27.0	0	.0886	.0723
28.0	5	.0976	.0779
29.0	5	.1067	.0837
30.0	6	.1175	.0896
31.0	10	.1356	.0958
32.0	0	.1356	.1022
33.0	3	.1410	.1088
34.0	3	.1465	.1156
35.0	0	.1465	.1226
36.0	4	.1537	.1298
37.0	0	.1537	.1372

AFQT percentile	Unsmoothed frequency	Cumulative proportion	
		Unsmoothed frequency	Smoothed frequency
38.0	4	.1609	.1448
39.0	0	.1609	.1527
40.0	4	.1682	.1607
41.0	0	.1682	.1690
42.0	3	.1736	.1774
43.0	0	.1736	.1861
44.0	4	.1808	.1949
45.0	0	.1808	.2040
46.0	6	.1917	.2133
47.0	0	.1917	.2227
48.0	3	.1971	.2324
49.0	5	.2061	.2423
50.0	11	.2260	.2524
51.0	0	.2260	.2627
52.0	10	.2441	.2732
53.0	0	.2441	.2840
54.0	7	.2568	.2949
55.0	0	.2568	.3060
56.0	8	.2712	.3173
57.0	0	.2712	.3289
58.0	8	.2857	.3406
59.0	13	.3092	.3526
60.0	0	.3092	.3648
61.0	17	.3400	.3771
62.0	0	.3400	.3897
63.0	8	.3544	.4025
64.0	0	.3544	.4155
65.0	14	.3797	.4287
66.0	14	.4051	.4421
67.0	0	.4051	.4557
68.0	15	.4322	.4695
69.0	0	.4322	.4835
70.0	15	.4592	.4977
71.0	0	.4593	.5121
72.0	15	.4864	.5268
73.0	0	.4864	.5416
74.0	22	.5262	.5567
75.0	0	.5262	.5719
76.0	18	.5588	.5874
77.0	0	.5588	.6031

AFQT percentile	Unsmoothed frequency	Cumulative proportion	
		Unsmoothed frequency	Smoothed frequency
78.0	19	.5931	.6189
79.0	0	.5931	.6350
80.0	15	.6203	.6513
81.0	0	.6203	.6678
82.0	28	.6709	.6845
83.0	17	.7016	.7014
84.0	0	.7016	.7185
85.0	16	.7306	.7359
86.0	15	.7577	.7534
87.0	16	.7866	.7711
88.0	21	.8246	.7891
89.0	0	.8246	.8072
90.0	20	.8608	.8255
91.0	17	.8915	.8441
92.0	0	.8915	.8629
93.0	20	.9277	.8818
94.0	0	.9277	.9010
95.0	10	.9458	.9204
96.0	0	.9458	.9400
97.0	15	.9729	.9598
98.0	9	.9892	.9798
99.0	6	1.0000	1.0000

TABLE E-2

CUMULATIVE DISTRIBUTION OF SAT SCORES
IN 1980 YOUTH SUBSAMPLE

SAT	Unsmoothed frequency	Cumulative proportion Unsmoothed frequency	Cumulative proportion Smoothed frequency
400	2	.0036	.0008
410	0	.0036	.0021
420	0	.0036	.0042
430	0	.0036	.0068
440	1	.0054	.0101
450	0	.0054	.0140
460	0	.0054	.0185
470	1	.0072	.0237
480	4	.0145	.0295
490	6	.0253	.0358
500	7	.0380	.0428
510	4	.0452	.0503
520	7	.0579	.0584
530	3	.0633	.0670
540	10	.0814	.0760
550	4	.0886	.0855
560	4	.0958	.0954
570	4	.1031	.1058
580	11	.1230	.1165
590	7	.1356	.1276
600	6	.1465	.1390
610	8	.1609	.1507
620	5	.1700	.1627
630	9	.1863	.1750
640	5	.1953	.1875
650	6	.2061	.2003
660	5	.2152	.2134
670	10	.2333	.2267
680	6	.2441	.2402
690	4	.2514	.2540
700	7	.2640	.2680
710	11	.2839	.2823
720	4	.2911	.2968
730	8	.3056	.3116
740	8	.3201	.3265
750	7	.3327	.3418
760	10	.3508	.3572

APPENDIX F

SCALING THE OFFICER SELECTION COMPOSITE

APPENDIX F

SCALING THE OFFICER SELECTION COMPOSITE

The score scale for the ASVAB subtests and aptitude composites, as well as AGCT, have traditionally been referenced to the WW II mobilization population. The aptitude composites derive their meaning through equated AFQT percentile scores. That is, the scale for the composite is built by equating its scores to AFQT percentiles. The actual scale value assigned to a particular sum of subtest standard scores (SSS) comprising the composite, is the score that would have been achieved by the WW II reference population with the associated AFQT percentile. A scale score value of 100 has the meaning of "average" (mean) expected performance, a score of 120 implies performance expected of the top 18 percent (100 - 82) percent of the reference population, and so on.

The Officer Selection Composite (OS) was scaled to place the scores on the same scale as the ASVAB-EL and AGCT. We also developed a scale (OS80) for the new composite based on the 1980 youth population [F-1], the new DoD reference population.

WWII Scale

The WW II, or traditional Army Standard Score Scale for the OS was constructed from the same data set used to scale the ASVAB 8/9/10. The sample data consisted of 5,375 enlisted applicants and recruits tested during 1980 [F-2].

Using this data set, we formulated the OS WWII scale in five steps.

1. Converting the four subtests comprising the OS to standard score form. The conversion from subtest raw, to standard score, uses a linear transformation of the form.

$$SS_{ij} = \frac{X_{ij} - \overline{X}_j}{SD} \; 10 + 50 \qquad \text{(F-1)}$$

where:

SS_{ij} = the _ith_ standard score on subtest j

X_{ij} = the _ith_ raw score on test j

$\overline{X}_j$, SD_j = the mean and standard deviation of the _jth_ subtest in the WWII reference population.

The means and standard deviations were estimated using a stratified, or weighted, distribution of the 5,375 scores in the combined applicant and recruit sample. The estimates are shown in table F-1. The weighting was done to simulate the distribution of scores in the reference population that had a uniform distribution of scores across deciles. The constants 10 and 50 were used to impose a standard deviation of 10 and a mean of 50 on the SS distributions.

TABLE F-1

SUBTEST MEANS AND STANDARD DEVIATIONS
IN STRATIFIED SAMPLE

ASVAB subtest	Mean	Standard deviation
GS	16.2	5.09
AR	17.8	2.70
AS	16.4	5.60
MK	12.5	5.95

2. Combining the four subtest standard scores into a sum of subtest standard scores (SSS), i.e.,

$$OS_{SSS} = GS_{SS} + MK_{SS} + AR_{SS} + AS_{SS} \ . \qquad (F-2)$$

The distributions of OS_{SSS} frequencies and cumulative proportions in both unsmoothed and smoothed form (using cubic splines) are shown in table F-2.

3. Computing AFQT-7A percentile scores, on the WWII scale, for the combined sample of applicants and recruits. The AFQT distribution is shown in table F-3.

4. Equipercentile equating of OS_{SSS} and AFQT scores, such that each SSS corresponds to a cumulative AFQT percent. For example, an SSS of 150 corresponds to an AFQT 11; 200 to 49; and 250 to 90. This equating can be interpreted to mean that 11 percent of the WWII reference population would have achieved an OS_{SSS} of up to 150; 49 percent an OS_{SSS} of 200; and so on. Figure F-1 is a graph of the cumulative OS_{SSS} and AFQT, used to do the equating, and table F-4 shows the OS_{SSS}.

TABLE F-2

CUMULATIVE DISTRIBUTION OF OS-SSS
IN ASVAB 8AX NORMALIZATION SUBSAMPLE

OS-SSS	Frequency	Cumulative proportion Unsmoothed	Cumulative proportion Smoothed
103.0	1	.0002	.0000
104.0	0	.0002	.0000
105.0	0	.0002	.0000
106.0	0	.0002	.0001
107.0	0	.0002	.0001
108.0	0	.0002	.0002
109.0	0	.0002	.0002
110.0	0	.0002	.0003
111.0	0	.0002	.0004
112.0	0	.0002	.0005
113.0	0	.0002	.0007
114.0	1	.0004	.0009
115.0	0	.0004	.0011
116.0	0	.0004	.0013
117.0	1	.0006	.0016
118.0	0	.0006	.0020
119.0	0	.0006	.0024
120.0	0	.0006	.0029
121.0	1	.0007	.0035
122.0	3	.0013	.0042
123.0	3	.0019	.0050
124.0	2	.0022	.0060
125.0	2	.0026	.0070
126.0	3	.0032	.0082
127.0	3	.0037	.0096
128.0	2	.0041	.0111
129.0	10	.0060	.0129
130.0	7	.0073	.0148
131.0	5	.0082	.0169
132.0	11	.0120	.0193
133.0	10	.0121	.0219
134.0	9	.0138	.0248
135.0	19	.0173	.0279
136.0	22	.0214	.0312
137.0	21	.0253	.0348
138.0	23	.0296	.0387
139.0	23	.0339	.0429

OS- SSS	Frequency	Cumulative proportion	
		Unsmoothed	Smoothed
140.0	18	.0372	.0473
141.0	25	.0419	.0520
142.0	28	.0471	.0570
143.0	38	.0541	.0622
144.0	30	.0597	.0677
145.0	33	.0659	.0735
146.0	32	.0718	.0795
147.0	38	.0789	.0858
148.0	28	.0841	.0923
149.0	39	.0913	.0991
150.0	37	.0982	.1062
151.0	42	.1060	.1135
152.0	37	.1129	.1210
153.0	38	.1200	.1287
154.0	43	.1280	.1367
155.0	45	.1364	.1449
156.0	43	.1444	.1533
157.0	53	.1542	.1620
158.0	44	.1624	.1708
159.0	45	.1708	.1798
160.0	66	.1831	.1890
161.0	52	.1927	.1983
162.0	49	.2019	.2078
163.0	59	.2128	.2175
164.0	48	.2218	.2273
165.0	67	.2342	.2372
166.0	55	.2445	.2472
167.0	53	.2543	.2573
168.0	48	.2633	.2675
169.0	48	.2722	.2779
170.0	69	.2850	.2883
171.0	60	.2962	.2987
172.0	68	.3088	.3093
173.0	63	.3206	.3198
174.0	53	.3304	.3305
175.0	50	.3397	.3412
176.0	51	.3492	.3519
177.0	50	.3585	.3627
178.0	60	.3697	.3735
179.0	52	.3793	.3844

OS- SSS	Frequency	Cumulative proportion	
		Unsmoothed	Smoothed
180.0	58	.3901	.3953
181.0	60	.4013	.4062
182.0	61	.4127	.4172
183.0	71	.4259	.4282
184.0	57	.4365	.4392
185.0	61	.4478	.4502
186.0	63	.4595	.4613
187.0	60	.4707	.4723
188.0	65	.4828	.4834
189.0	64	.4947	.4945
190.0	58	.5055	.5055
191.0	61	.5168	.5165
192.0	59	.5278	.5275
193.0	49	.5369	.5385
194.0	55	.5472	.5494
195.0	68	.5598	.5603
196.0	56	.5702	.5712
197.0	58	.5810	.5820
198.0	46	.5896	.5927
199.0	64	.6015	.6034
200.0	99	.6199	.6140
201.0	47	.6287	.6245
202.0	62	.6402	.6349
203.0	47	.6489	.6453
204.0	53	.6588	.6555
205.0	51	.6683	.6656
206.0	55	.6785	.6756
207.0	56	.6889	.6854
208.0	40	.6964	.6952
209.0	56	.7068	.7048
210.0	44	.7150	.7142
211.0	60	.7261	.7236
212.0	63	.7379	.7328
213.0	43	.7459	.7418
214.0	51	.7553	.7507
215.0	50	.7647	.7594
216.0	46	.7732	.7680
217.0	33	.7793	.7764
218.0	50	.7887	.7846
219.0	49	.7978	.7927

OS-SSS	Frequency	Cumulative proportion	
		Unsmoothed	Smoothed
220.0	42	.8056	.8006
221.0	50	.8149	.8083
222.0	43	.8229	.8159
223.0	37	.8298	.8232
224.0	40	.8372	.8304
225.0	35	.8437	.8374
226.0	24	.8482	.8443
227.0	35	.8547	.8510
228.0	36	.8614	.8575
229.0	36	.8681	.8639
230.0	28	.8733	.8701
231.0	28	.8785	.8761
232.0	30	.8841	.8820
233.0	32	.8900	.8877
234.0	49	.8992	.8933
235.0	24	.9036	.8988
236.0	22	.9077	.9041
237.0	28	.9127	.9092
238.0	31	.9187	.9143
239.0	26	.9235	.9191
240.0	22	.9276	.9239
241.0	25	.9323	.9285
242.0	28	.9375	.9329
243.0	16	.9405	.9372
244.0	16	.9434	.9414
245.0	24	.9479	.9455
246.0	23	.9522	.9494
247.0	18	.9555	.9532
248.0	13	.9580	.9569
249.0	23	.9622	.9604
250.0	16	.9652	.9638
251.0	25	.9699	.9671
252.0	14	.9725	.9702
253.0	16	.9754	.9732
254.0	13	.9779	.9760
255.0	14	.9805	.9787
256.0	21	.9844	.9813
257.0	18	.9877	.9837
258.0	10	.9896	.9859
259.0	8	.9911	.9880

OS-SSS	Frequency	Cumulative proportion	
		Unsmoothed	Smoothed
260.0	10	.9929	.9899
261.0	7	.9942	.9917
262.0	5	.9952	.9933
263.0	10	.9970	.9947
264.0	4	.9978	.9960
265.0	8	.9993	.9971
266.0	1	.9994	.9980
267.0	1	.9996	.9988
268.0	1	.9998	.9993
269.0	0	.9998	.9998
270.0	1	1.0000	1.0000

TABLE F-3

CUMULATIVE DISTRIBUTION OF AFQT 7A PERCENTILE SCORES
IN ASVAB 8AX NORMALIZATION SUBSAMPLE

AFQT percentile	Unsmoothed frequency	Cumulative proportion	
		Unsmoothed	Smoothed
1.0	27	.0050	.0043
2.0	13	.0074	.0096
3.0	17	.0106	.0159
4.0	14	.0132	.0232
5.0	46	.0218	.0315
6.0	26	.0266	.0408
7.0	45	.0350	.0511
8.0	35	.0415	.0623
9.0	87	.0577	.0744
10.0	50	.0670	.0874
11.0	114	.0882	.1012
12.0	73	.1018	.1157
13.0	111	.1224	.1308
14.0	70	.1354	.1465
15.0	144	.1622	.1626
16.0	106	.1820	.1791
17.0	165	.2127	.1959
18.0	101	.2314	.2127
19.0	105	.2510	.2297
20.0	76	.2651	.2466
21.0	71	.2783	.2634
22.0	30	.2839	.2801
23.0	93	.3012	.2966
24.0	108	.3213	.3128
25.0	102	.3403	.3288
26.0	28	.3455	.3444
27.0	112	.3663	.3598
28.0	93	.3836	.3747
29.0	89	.4002	.3893
30.0	24	.4047	.4036
31.0	109	.4249	.4174
32.0	110	.4454	.4308
33.0	102	.4644	.4437
34.0	31	.4701	.4563
35.0	0	.4701	.4685
36.0	116	.4917	.4804

AFQT percentile	Unsmoothed frequency	Cumulative proportion	
		Unsmoothed	Smoothed
37.0	0	.4917	.4919
38.0	116	.5133	.5032
39.0	0	.5133	.5142
40.0	101	.5321	.5250
41.0	0	.5321	.5356
42.0	32	.5380	.5461
43.0	0	.5380	.5566
44.0	108	.5581	.5670
45.0	0	.5581	.5774
46.0	104	.5775	.5879
47.0	0	.5775	.5984
48.0	118	.5994	.6089
49.0	28	.6047	.6195
50.0	113	.6257	.6300
51.0	133	.6504	.6406
52.0	106	.6701	.6511
53.0	0	.6701	.6616
54.0	21	.6740	.6720
55.0	0	.6740	.6823
56.0	112	.6949	.6926
57.0	0	.6949	.7027
58.0	104	.7142	.7128
59.0	0	.7142	.7228
60.0	107	.7341	.7326
61.0	13	.7366	.7423
62.0	126	.7600	.7519
63.0	99	.7784	.7613
64.0	0	.7784	.7705
65.0	102	.7974	.7796
66.0	0	.7974	.7885
67.0	24	.8019	.7972
68.0	0	.8019	.8057
69.0	0	.8019	.8142
70.0	111	.8225	.8225
71.0	1	.8227	.8307
72.0	100	.8413	.8389
73.0	0	.8413	.8470
74.0	96	.8592	.8550
75.0	1	.8593	.8629

AFQT percentile	Unsmoothed frequency	Cumulative proportion	
		Unsmoothed	Smoothed
76.0	15	.8621	.8708
77.0	0	.8621	.8787
78.0	96	.8800	.8865
79.0	0	.8800	.8943
80.0	87	.8962	.9019
81.0	90	.9129	.9095
82.0	10	.9148	.9170
83.0	74	.9286	.9243
84.0	0	.9286	.9314
85.0	66	.9408	.9383
86.0	0	.9408	.9450
87.0	63	.9526	.9514
88.0	7	.9539	.9576
89.0	62	.9654	.9634
90.0	52	.9751	.9689
91.0	42	.9829	.9740
92.0	5	.9838	.9788
93.0	33	.9900	.9831
94.0	18	.9933	.9870
95.0	10	.9952	.9905
96.0	2	.9955	.9936
97.0	14	.9981	.9961
98.0	8	.9996	.9983
99.0	2	1.0000	1.0000

TABLE F-4

EQUIVALENT AFQT 7A AND OS-SSS
IN ASVAB 8AX NORMALIZATION SAMPLE

AFQT percentile	Corresponding scale score	OS-SSS	AFQT percentile	Corresponding scale score	OS-SSS
1	40	123	18	77	163
1	40	124	18	77	164
1	40	125	19	78	165
1	40	126	20	79	166
2	45	127	20	79	167
2	45	128	21	80	168
2	45	129	21	80	169
2	45	130	22	81	170
3	50	131	23	82	171
3	50	132	23	82	172
3	50	133	24	82	173
4	54	134	25	84	174
4	54	135	25	84	175
4	54	136	26	85	176
5	57	137	27	86	177
5	57	138	27	86	178
6	59	139	28	87	179
6	59	140	29	88	180
7	61	141	30	89	181
7	61	142	30	89	182
7	61	143	31	90	183
8	62	144	32	91	184
8	62	145	33	92	185
9	64	146	34	92	186
9	64	147	35	93	187
10	65	148	36	93	188
10	65	149	37	94	189
11	66	150	38	94	190
11	66	151	39	95	191
12	68	152	40	95	192
12	68	153	41	96	193
13	70	154	42	96	194
13	70	155	43	97	195
14	71	156	44	97	196
14	71	157	45	98	197
15	74	158	46	98	198
16	75	159	47	99	199
16	75	160	48	99	200
17	76	161	49	100	201
17	76	162	50	100	202

AFQT percentile	Corresponding scale score	OS-SSS	AFQT percentile	Corresponding scale score	OS-SSS
51	101	203	84	121	243
52	102	204	85	122	244
53	103	205	86	123	245
54	103	206	86	123	246
55	104	207	87	123	247
56	104	208	87	123	248
57	105	209	88	124	249
58	105	210	89	125	250
59	106	211	89	125	251
60	107	212	90	126	252
60	107	213	90	126	253
61	107	214	91	127	254
62	108	215	91	127	255
63	109	216	92	128	256
64	109	217	93	130	257
65	110	218	93	130	258
66	111	219	94	131	259
67	111	220	94	131	260
68	112	221	95	133	261
69	112	222	95	133	262
70	113	223	96	135	263
70	113	224	96	135	264
71	113	225	97	137	265
72	114	226	97	137	266
73	114	227	98	141	267
74	115	228	98	141	268
75	116	229	98	141	269
75	116	230	98	141	270
76	116	231			
77	117	232			
78	117	233			
78	117	234			
79	118	235			
80	118	236			
80	118	237			
81	119	238			
82	120	239			
83	120	240			
82	121	241			
84	121	242			

5. Translating the percentiles associated with each OS_{SSS} to an aptitude composite standard scale score. This last step is simply a "table look-up." This table, shown on table F-5 relates AFQT percentiles to aptitude composite standard scale scores in the WWII reference population [Maier and Grafton]. The final aptitude score scale has a mean of 100 and a standard deviation of 20, and is on the same score scale used for ASVAB-EL and AGCT. Table F-6 is the conversion table for translating OS_{SSS} to its equivalent aptitude composite score on the WWII scale. Also shown in table F-6, is the conversion of OS_{SSS} to 1980 aptitude composite scores. The next section describes how the 1980 aptitude composite scores were derived.
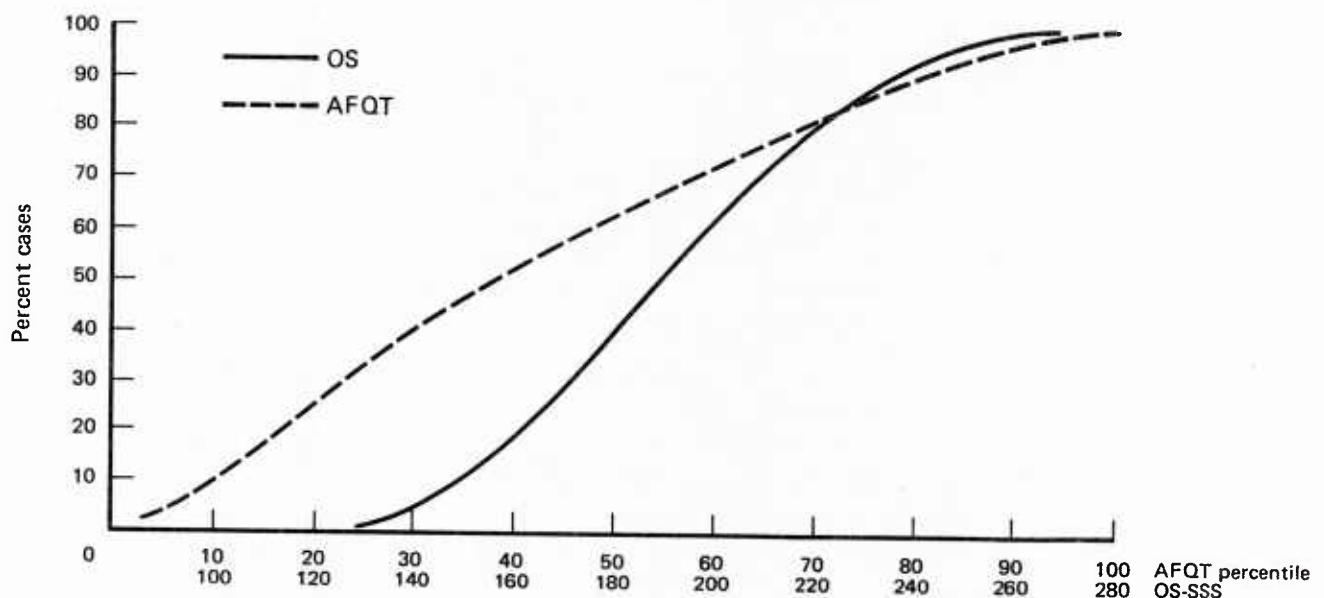

FIG. F-1: EQUIPERCENTILE EQUATING OF AFQT PERCENTILE AND OFFICER SELECTION COMPOSITE SUM OF SUBTEST STANDARD SCORES IN SAMPLE OF 5,375 APPLICANTS AND RECRUITS

1980 Score Scale

The 1980 scale was constructed using the 1980 youth sample of 18- to 23-year-olds. The method consisted of the following steps. First, compute the subtest standard scores with equation (F-1), using the 1980 reference population means and standard deviations shown in table F-7, for the linear transformations. Then, sum the subtest standard scores to form the OS_{SSS}. The last step is to convert the SSS to the 1980 score scale using the simple linear transformation shown in equation (F-3). Note that no intermediate step using AFQT percentiles is used. The scaled composite is based directly on the distribution of the OS_{SSS} in the 1980 population.

$$OS80 = \frac{OS_{SSS} - \mu}{\sigma} \ 20 + 100 \ , \tag{3}$$

where μ and σ are the mean and standard deviation of the OS_{SSS} in the 1980 reference population. The constants, 20 and 100, are used to impose a standard deviation of 20 and mean of 100 on the OS80 score scale. The OS_{SSS} to OS80 conversions are shown in table F-6.

Equivalent OS scores on the 1980 and WWII Score Scales

The final analysis performed was the equipercentile equating of the 1980 and WWII Score Scales for the OS composite. The 1980 youth sample was used for this purpose.

The equating procedure consisted of the following steps. We computed subtest standard scores for each person in the 1980 youth sample, for each scale, with equation F-1, substituting the means and standard deviations for the appropriate reference population (table F-1 for the WWII and F-7 for 1980). The subtest standard scores were then summed. Then the OS_{SSS} were converted to aptitude composite scale score form, using table F-6. The distributions of OS WWII scores is shown in table F-8, and OS80 in table F-9. The smoothed frequencies were then accumulated and converted to cumulative proportions.

Figure F-2 shows the cumulative distributions of the OS composite based on the 1980 and World War II scale scores. These data indicate that an OS of 120 on the WW II scale is equivalent to a value of 124 on the 1980 scale. Table F-10 lists the equivalent OS scores on the two scales.

TABLE F-5

CONVERSION TABLE OF
PERCENTILE SCORES TO ARMY STANDARD SCORES

Percentile score	Army standard score	Percentile score	Army standard score	Percentile score	Army standard score
100	155	65	110	30	89
99	147	64	109	29	88
98	141	63	109	28	87
97	137	62	108	27	86
96	135	61	107	26	85
95	133	60	107	25	84
94	131	59	106	24	83
93	130	58	105	23	82
92	128	57	105	22	81
91	127	56	104	21	80
90	126	55	104	20	79
89	125	54	103	19	78
88	124	53	103	18	77
87	123	52	102	17	76
86	123	51	101	16	75
85	122	50	100	15	74
84	121	49	100	14	71
83	121	48	99	13	70
82	120	47	99	12	68
81	119	46	98	11	66
80	118	45	98	10	65
79	118	44	97	9	64
78	117	43	97	8	62
77	117	42	96	7	61
76	116	41	96	6	59
75	116	40	95	5	57
74	115	39	95	4	54
73	114	38	94	3	50
72	114	37	94	2	45
71	113	36	93	1	40
70	113	35	93		
69	112	34	92		
68	112	33	92		
67	111	32	91		
66	111	31	90		

SSS	1980	WWII
254	132	127
255	133	128
256	133	129
257	134	130
258	135	131
259	135	131
260	136	132
261	136	133
262	137	134
263	137	135
264	138	136
265	139	137
266	139	138
267	140	140
268	140	141
269	141	141
270	142	141
271	142	141
272	142	141

TABLE F-7

SUBTEST MEANS AND STANDARD DEVIATIONS
IN 1980 YOUTH REFERENCE POPULATION

ASVAB subtest	Mean	Standard deviation
GS	15.9	5.01
AR	18.0	7.37
AS	14.3	5.55
MK	13.6	6.39

CUMULATIVE DISTRIBUTION OF OS WWII SCALE SCORES
IN THE 1980 POPULATION

Scale score	Frequency	Cumulative proportion	
		Unsmoothed	Smoothed
40	200,815	.0079	.0004
41	0	.0079	.0010
42	0	.0079	.0018
43	0	.0079	.0028
44	0	.0079	.0041
45	137,611	.0133	.0055
46	0	.0133	.0072
47	0	.0133	.0090
48	0	.0133	.0112
49	0	.0133	.0135
50	161,203	.0197	.0161
51	0	.0197	.0189
52	0	.0197	.0220
53	0	.0197	.0254
54	223,658	.0285	.0290
55	0	.0285	.0329
56	0	.0285	.0371
57	175,686	.0354	.0417
58	0	.0354	.0465
59	205,650	.0435	.0516
60	0	.0435	.0571
61	354,705	.0574	.0628
62	220,047	.0661	.0689
63	0	.0661	.0753
63	293,629	.0776	.0821
65	325,882	.0905	.0891
66	304,145	.1024	.0965
67	0	.1024	.1043
68	373,517	.1171	.1123
69	0	.1171	.1207
70	352,645	.1310	.1294
71	345,104	.1446	.1384
72	0	.1446	.1478
73	0	.1446	.1575
74	177,469	.1516	.1676
75	342,426	.1651	.1780

Scale score	Frequency	Cumulative proportion	
		Unsmoothed	Smoothed
76	331,530	.1781	.1888
77	338,245	.1914	.1999
78	205,868	.1995	.2113
79	395,784	.2151	.2231
80	425,272	.2318	.2352
81	162,262	.2382	.2477
82	771,339	.2686	.2605
83	0	.2686	.2736
84	407,675	.2846	.2870
85	225,056	.2935	.3007
86	497,236	.3131	.3147
87	259,969	.3233	.3291
88	249,608	.3331	.3437
89	483,951	.3522	.3585
90	205,737	.3603	.3737
91	178,193	.3673	.3891
92	536,201	.3884	.4048
93	423,828	.4051	.4207
94	485,439	.4242	.4368
95	423,460	.4408	.4531
96	435,208	.4579	.4697
97	494,899	.4774	.4863
98	529,208	.4983	.5031
99	725,400	.5268	.5201
100	483,237	.5458	.5371
101	275,382	.5567	.5542
102	187,530	.5640	.5714
103	534,353	.5851	.5885
104	460,841	.6032	.6057
105	461,683	.6214	.6229
106	189,121	.6288	.6400
107	694,323	.6561	.6571
108	211,682	.6645	.6740
109	437,100	.6817	.6909
110	160,290	.6880	.7076
111	492,377	.7074	.7241
112	495,641	.7269	.7405
113	721,791	.7553	.7566
114	360,682	.7695	.7724
115	219,031	.7781	.7879

Scale score	Frequency	Cumulative proportion Unsmoothed	Cumulative proportion Smoothed
116	600,185	.8017	.8032
117	630,657	.8273	.8180
118	615,762	.8516	.8324
119	261,440	.8618	.8464
120	279,854	.8729	.8599
121	537,035	.8940	.8729
122	194,443	.9016	.8853
123	600,462	.9253	.8972
124	121,111	.9300	.9085
125	308,415	.9422	.9192
126	293,232	.9537	.9292
127	225,132	.9626	.9386
128	108,040	.9668	.9474
129	0	.9668	.9555
130	182,879	.9740	.9629
131	159,950	.9803	.9697
132	0	.9803	.9757
133	213,746	.9887	.9811
134	0	.9887	.9859
135	109,141	.9930	.9899
136	0	.9930	.9933
137	100,094	.9970	.9960
138	0	.9970	.9980
139	0	.9970	.9993
140	0	.9970	1.0000
141	76,820	1.0000	1.0000

CUMULATIVE DISTRIBUTION OF OS 1980 SCALE SCORE
IN THE 1980 POPULATION

Scale score	Frequency	Cumulative proportion	
		Unsmoothed	Smoothed
40	20,702	.0008	.0000
41	0	.0008	.0001
42	5,181	.0010	.0004
43	0	.0010	.0008
44	3,138	.0011	.0013
45	1,026	.0012	.0019
46	1,667	.0012	.0027
47	922	.0013	.0037
48	3,316	.0014	.0047
49	1,121	.0015	.0060
50	2,318	.0016	.0074
51	3,469	.0017	.0091
52	4,873	.0019	.0109
53	9,994	.0023	.0129
54	10,483	.0027	.0152
55	12,492	.0032	.0178
56	22,954	.0041	.0206
57	14,637	.0047	.0237
58	52,809	.0067	.0271
59	22,965	.0076	.0309
60	69,314	.0104	.0350
61	72,477	.0132	.0395
62	49,131	.0152	.0444
63	114,465	.0197	.0497
64	157,410	.0259	.0554
65	79,771	.0290	.0615
66	161,971	.0354	.0681
67	215,776	.0439	.0751
68	228,581	.0529	.0825
69	125,811	.0578	.0904
70	214,885	.0663	.0987
71	288,685	.0776	.1075
72	161,321	.0840	.1167
73	328,597	.0969	.1264
74	332,553	.1100	.1365
75	392,590	.1254	.1470

Scale score	Frequency	Cumulative proportion	
		Unsmoothed	Smoothed
76	172,084	.1322	.1578
77	318,801	.1448	.1691
78	372,752	.1594	.1808
79	154,700	.1655	.1928
80	346,331	.1792	.2052
81	351,834	.1930	.2179
82	410,490	.2092	.2310
83	185,783	.2165	.2443
84	431,075	.2334	.2580
85	401,169	.2492	.2720
86	279,315	.2602	.2862
87	460,812	.2783	.3007
88	508,584	.2984	.3155
89	454,418	.3162	.3304
90	294,197	.3278	.3456
91	473,565	.3465	.3610
92	435,153	.3636	.3765
93	235,251	.3728	.3921
94	527,105	.3936	.4079
95	373,370	.4083	.4238
96	540,099	.4295	.4398
97	224,518	.4384	.4559
98	445,740	.4559	.4721
99	493,698	.4754	.4883
100	269,163	.4859	.5045
101	741,759	.5151	.5207
102	493,697	.5346	.5370
103	249,313	.5444	.5531
104	533,491	.5654	.5693
105	448,870	.5830	.5854
106	532,963	.6040	.6013
107	211,950	.6124	.6172
108	516,658	.6327	.6330
109	444,624	.6502	.6486
110	194,582	.6578	.6640
111	468,014	.6763	.6793
112	394,753	.6918	.6945
113	428,712	.7087	.7094
114	338,398	.7220	.7241
115	491,696	.7413	.7387

| Scale score	Frequency	Cumulative proportion	
		Unsmoothed	Smoothed
116	447,441	.7590	.7530
117	176,466	.7659	.7670
118	397,872	.7816	.7808
119	438,511	.7988	.7943
120	203,065	.8068	.8076
121	213,642	.8152	.8205
122	448,086	.8329	.8332
123	414,097	.8492	.8455
124	282,086	.8603	.8575
125	292,438	.8718	.8692
126	351,152	.8856	.8805
127	390,582	.9010	.8914
128	161,400	.9073	.9019
129	341,908	.9208	.9120
130	233,522	.9300	.9216
131	158,798	.9362	.9308
132	318,689	.9487	.9396
133	278,442	.9597	.9479
134	70,347	.9625	.9557
135	228,306	.9715	.9630
136	161,545	.9778	.9698
137	222,187	.9866	.9761
138	49,180	.9885	.9819
139	115,453	.9930	.9872
140	111,760	.9974	.9920
141	28,841	.9986	.9963
142	36,314	1.0000	1.0000
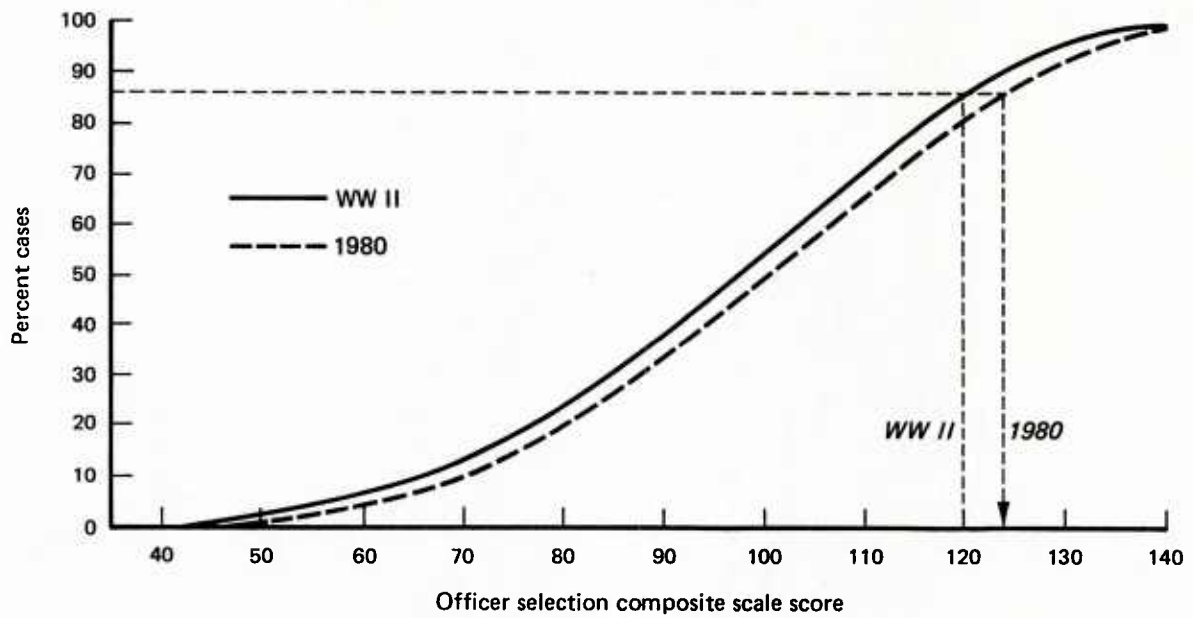
FIG. F-2: EQUIPERCENTILE EQUATING OF OS COMPOSITE SCORES SCALED
TO 1980 DoD AND WORLD WAR II REFERENCE POPULATIONS
IN 1980 YOUTH SAMPLE

TABLE F-10

EQUIVALENT ASVAB-OS SCORES
IN THE 1980 WWII POPULATION

OS score		OS score	
WW II	1980	WW II	1980
40	42	80	82
41	43	81	83
42	45	82	84
43	46	83	85
44	47	84	86
45	49	85	87
46	50	86	88
47	51	87	89
48	52	88	90
49	53	89	91
50	54	90	92
51	55	91	93
52	56	92	94
53	57	93	95
54	58	94	96
55	59	95	97
56	60	96	98
57	61	97	99
58	62	98	100
59	63	99	101
60	64	100	102
61	65	101	103
62	66	102	104
63	67	103	105
64	68	104	106
65	69	105	107
66	70	106	108
67	71	107	110
68	72	108	111
69	72	109	112
70	73	110	113
71	74	111	114
72	75	112	115
73	76	113	116
74	77	114	117
75	78	115	119
76	79	116	120
77	80	117	121
78	80	118	122
79	81	119	123

OS Score	
WW II	1980
120	124
121	125
122	126
123	128
124	129
125	130
126	131
127	132
128	133
129	134
130	135
131	136
132	137
133	138
134	139
135	140
136	140
137	141
138	141
139	142